SecureLearn: Aplicação de Machine Learning no contexto da Lei Geral de Proteção de Dados (LGPD)

Juan Cardoso da Silva, Ronaldo Celso Messias Correia

Departamento de Matemática e Computação (DMC)

Universidade Estadual Paulista "Júlio Mesquita Filho" (UNESP)

Faculdade de Ciências e Tecnologia (FCT)

Presidente Prudente - SP, Brasil

juan.c.silva@unesp.br, ronaldo.correia@unesp.br

Resumo. *Introduzidas as leis protetoras privacidades no mundo digital ao redor do mundo como a LGPD (Lei Geral de Proteção de Dados) – Surgiu um problema relacionado a proteção de dados privados onde sistemas não têm mecanismos adequados para o segurança dos dados. Neste trabalho é demonstrado como modelos de aprendizado de máquina supervisionados são capazes de identificar os dados, julgar o contexto dos dados e enviar os dados para um algoritmo de encriptação. No artigo é explorado como um algoritmo inteligente realiza o processo de garantir o cumprimento da lei, a quantidade de detecções e a capacidade de julgar dados a partir de uma base de dados gerados, mostrando as deteções variadas distribuídas entre verdadeiros positivos, negativos e falsos positivos, negativos – demonstrando a capacidade da área de inteligência artificial em ajudar a área de direito e complementar a área de segurança da informação.*

Palavras Chaves: Lei Geral de Proteção de dados (LGPD); Métodos de aplicação da LGPD; Machine Learning; Inteligência Artificial; Tensorflow.

SecureLearn: Machine Learning application in the contexto of General Data Protection Law (LGPD)

Abstract. *With the introduction of data protection regulations around the digital world such as LGPD (General Data Protection Law) – A problem has arisen regarding the protection of private data where systems do not have adequate mechanisms for data security. This paper demonstrates how supervised machine learning models are able to identify data, judge the context of the data and send the data to an encryption algorithm. The article explores how an intelligent algorithm carries out the process of ensuring compliance with the law, the number of detections and the ability to judge data from a generated database, showing the varied detections distributed between true positives, negatives and false positives, negatives - demonstrating the ability of the area of artificial intelligence to help the area of law and complement the area of information security.*

Keywords: General Data Protection Law (LGPD); Methods of applying LGPD; Machine Learning; Artificial Inteligence; Tensorflow**.**

---

1. INTRODUÇÃO

Com a introdução da *General Data Protection Regulation* (GDPR)[1], o judiciário brasileiro implementou uma lei irmã nomeada de Lei Geral de Proteção de dados (LGPD), com o foco garantir os direitos do cidadão brasileiro no mundo digital, como a sua privacidade garantida pela constituição – a lei engloba os crimes digitais cometidos por ataques hackers, como também desleixos resultantes de vazamento de dados privados no ambiente de trabalho, existem outros direitos garantidos pela lei, mas o foco neste projeto é o artigo 6 da LGPD.

O artigo tem como função guiar as pessoas responsáveis pelo tratamento de dados em boa-fé para garantir os direitos constitucionais de privacidade no mundo digital, seja em sistemas, serviços online, acessando redes sociais e outros web serviços. Os modelos criados respeitam os incisos I (tratar os dados legitimamente), VII (meios para proteger os dados), VIII (evitar danos aos dados utilizados em questão) e IX (tratar os dados sem descriminação), o motivo dos incisos específicos e não o artigo inteiro é por questão de experimentação, os modelos desenvolvidos neste projeto realizam a identificação de dados, retorna um vetor de contendo as previsões para um algoritmo de encriptação, essas operações feitas se encaixam no contexto desses incisos do artigo, invés dele como um todo.

A maioria dos sistemas atuais conectados na internet nem sempre possuem os melhores mecanismos de segurança quando se tratam de dados digitais ou muitas vezes acabam ocorrendo desleixos nos sistemas permitindo o acesso de dados privados/sensíveis de usuários a caírem em mãos de agentes mal intencionados como hackers e golpistas utilizando *ransomware* (encriptação ilegal de dados e estorno destes a pedido de dinheiro). Outro problema é a falta de restrição de acesso a dados quando estes são disponibilizados por interfaces de visualizações de dados tais como o PowerBi da Microsoft, resultando na confiabilidade da empresa no indivíduo em questão com acesso aos dados para evitar desleixo ao mexer com os dados como utilizá-los em segurança.

Uso de inteligência artificial (IA) para garantir integridade de dados e cumprir a lei está sendo estudado como são os casos da IBM, onde utiliza o QRadar[2] em conjunto com um modelo de machine learning para detectar falhas/vulnerabilidades do sistema e o modelo próprio focado em cortar dados irrelevantes. Esses métodos garantem partes da aplicação da lei, mas não garante a segurança, QRadar apenas avisa sobre falhas e o modelo em apenas corta dados bem específicos, como mostrar salários de até um alcance determinado pela lei.

[1]A lei europeia pode ser encontrada em: https://gdpr-info.eu

[2]O artigo do modelo QRadar pode ser encontrado em: https://www.proof.com.br/wp-content/uploads/2019/08/Using-QRadar-for-LGPD.pdf

O modelo QRadar possui uma limitação óbvia em relação em sua aplicação sendo a capacidade de realizar outras operações além de avisar, já que apenas garantir um aviso a um administrador não garante a integridade e sim os requisitos pela lei sobre o aviso ao usuário dono dos dados sobre a vulnerabilidade – o modelo de corte da IBM[3] realiza os cortes dos dados sensíveis quando pedido, mas não garante a segurança.

Neste artigo é apresentado como a LGPD pode ser utilizada em conjunto de modelos de machine learning, identificando os dados relevantes e encriptando eles, deixando os dados irrelevantes sem encriptação, por fim, comparar os resultados dos modelos a apresentados, suas limitações e assegurar os incisos mencionados sejam respeitados.

Na seção II Metodologias é mostrado como foram desenvolvido os modelos para solucionar o problema apresentado, na seção III é apresentado o estudo de caso como o modelo se comporta para solucionar o problema, em IV é apresentado o desempenho dos modelos, V é a seção de limitações e discussões de como alguns dos processos apresentados na seção II pode ter impactado o projeto, em VI as possíveis aplicações de um projeto como este, na seção VII os trabalhos futuros e por fim as conclusões tiradas após a experimentação e desenvolvimento do projeto.

2. REVISÃO TEÓRICA

Esta seção é dedicada a discussão sobre os conceitos básicos relacionados aos temas da LGPD, Machine Learning, encriptação e ferramentas de análise para os resultados.

2.1 LEI GERAL DE PROTEÇÃO DE DADOS

Criada em 2018 e entrado em vigor em 2021, a Lei Geral de Proteção de Dados 13.709/2018 (abreviando como LGPD) possui o objetivo de garantir os direitos de privacidade e proteção de dados sensíveis de usuários e o livre desenvolvimento da personalidade pessoal natural – a lei atua também como um mecanismo de segurança judiciária, promovendo padronizações, regulamentos, práticas de proteção de dados pessoais/sensíveis e a punição por multas e até prisões, caso ocorra não realização da lei.

Pela lei, dados sensíveis/privados são informações relacionadas diretamente à intimidade pessoal de um indivíduo, opiniões políticas, convicções religiosas, filiação a organizações religiosas/políticas/filosóficas, etnia, dados relacionados à saúde, dados genéticos (como biometria e afins), CPF e todo outro tipo de informação do qual possa ser vinculada a vida íntima de uma pessoa.

[3]O artigo do modelo apresentado pode ser encontrado em: https://link.springer.com/article/10.1007/s43681-021-00095-8

A lei garante ao cidadão a capacidade de deleção dos dados sensíveis/privados, revogar consentimento do uso de tais dados, transferir os dados para outro serviço similar ao serviço atual, acesso aos dados para atualizá-los e garantir a consistência dos dados no meio digital.

Além disso a lei, também dispõe significados para como esses dados podem ser manipulados, pela lei, o tratamento de dados é entendido como qualquer tipo de operação onde os dados pessoais recebem interação por parte de um profissional, tais como acessar um dado, alterar, deletar, produzir, coletar, classificar, acessar, reproduzir, copiar, processar, armazenar e quais queres outras manipulações possíveis com os dados.

A lei também determina a existência de um operador e um controlador, operadores são pessoas das quais possam tomar decisões finais a respeito de como o dado vai ser manipulado e os operadores realizam as manipulações, ambos podem ser pessoas naturais ou jurídicas, sendo elas do âmbito público ou privado.

Pelas considerações do Art. 6[4], o tratamento de dados pessoais deve seguir alguns princípios, para este projeto, os princípios considerados para atingir o objetivo procurado são transparência, adequação, finalidade e responsabilização.

Os princípios seguem como:

I. **Finalidade:** realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
II. **Adequação:** compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;
III. **Necessidade:** limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;
IV. **Livre Acesso:** garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;
V. **Qualidade de dados:** garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

[4]Pode ser acessado para consulta sobre em: https://lgpd-brasil.info/capitulo_01/artigo_06

VI. **Transparência:** garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

VII. **Segurança:** utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;

VIII. **Prevenção:** adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;

IX. **Não discriminação:** impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

X. **Responsabilização:** demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

2.2 MACHINE LEARNING

Machine Learning, uma categoria pertencente à área de IA tem como objetivo "treinar" máquinas com intenção de resolver problemas, automatizar tarefas e identificar padrões.

Com as capacidades modernas da *Machine Learning* aumentando, a área deixou de apenas ser estruturação e dedução de dados, evoluindo para entender padrões indefinidos usando técnicas de treinamento do qual cada padrão aprendido utilizado, pode ser utilizado para analisar, outros padrões indefinidos, compreender dados e automatizar tarefas. Hoje me dia, com a capacidade computacional da nuvem e a abrangente quantidade de dados disponíveis, ou até mesmo gerados por outras ML para treinar ML, demonstrando a capacidade de ser aplicada em diversos cenários diferentes para chegar a um resultado esperado/procurado, ajudando em várias áreas de atuações nas profissões, com as maiores contribuições da ML sendo para as áreas da saúde e áreas de processamento de imagens. Este trabalho foca em trazer essas ideias de aplicações de ML para área de direito aplicada, mais especificamente na aplicação de LGPD.

Utilizando *Deep Learning* podemos usar a ML para aprender a não só reconhecer e identificar padrões, como utilizar para realizar avaliações de dados, seguindo algumas categorias de avaliações.

2.3 TENSORFLOW

TensorFlow[5], é um framework de algoritmos de *Deep Learning* e *Machine Learning*, desenvolvido pela *Google Brains* e tornado *open-source* em 2015, sendo uma biblioteca aberta utilizando computação numérica em larga escala

O *TensorFlow* cria grafos/estruturas permitindo o fluxo destes sejam processados, cada vértice do grafo possui a representatividade de operações matemáticas e cada aresta, um array multidimensional, chamado de *tensor*.

A biblioteca permite funcionar localmente, consumindo GPU e CPU da máquina do programador ou utilizando uma máquina na nuvem, com um maior poder computacional, neste caso, utilizando uma TPU resultando em maior poder computacional. No desenvolvimento do projeto, no entanto, foi utilizado GPU em conjunto da CPU localmente em um Tensorflow instalado localmente na máquina e rodando dentro de um ambiente curado no Jupyter-Notebook onde foram realizados tratamento de dados, criação de modelos e testes.

2.4 JUPYTER NOTEBOOK

Jupyter-Notebook é um ambiente para executar e projetar códigos na web utilizando python como sua linguagem de programação, cada notebook é dividido em seções acompanhadas com áreas de textos contendo markdown text para anotações e títulos. Cada seção dentro de um notebook (chamada de documentos) possui um input e um output, servindo como um espaço de log para cada documento e permitir verificar resultados anteriores como permitir uma modularidade maior para corrigir e escrever código.

2.5 MATPLOTLIB, NUMPY E PANDSA

Numpy é uma biblioteca de Python criada utilizando vetorização invés de loops, arrays e indexação, utilizando código de máquina pré compilado em C baseada no objeto *array* da biblioteca para realizar as operações com velocidade e consistência, pois todas as operações necessárias estão pré-compiladas e armazenadas pronto para serem utilizadas.

Pandas[6] permite o cientista poder manipular os dados sem necessitar preocupar com tratamento das estruturas, complexidade e tempo de velocidade dentro de um algoritmo, facilitando a operações/manipulações com dados, estruturando-os para uso em conjunto de outras aplicações, como aprendizado de máquina, visualização de dados ou até mesmo Cálculo Numérico.

[5]Metodologia de uso do tensorflow para esse projeto pode ser acessada em: https://ieeexplore.ieee.org/document/8093521

[6]Guias utilizado para aprender a usar o Pandas disponível em: https://pandas.pydata.org/about/

Matplotlib[7] é uma biblioteca implementada em Python para visualização de dados em forma gráfica, permitindo visualização de dados conforme a necessidade do programador ao realizar testes, experimentos, pesquisas e trabalhar com resultados obtidos de projetos.

2.6 CRIPTOGRAFIA

RSA é um algoritmo de criptografia desenvolvido em 1976 logo depois do primeiro algoritmo de compartilhamento de chaves chamado Diffie-Hellman apresentado. O RSA permite a encriptação de dados e a desencriptarão dos mesmos utilizando uma chave pública, privada e um números primos para geração dessas chaves.

Primeiro selecionamos um valor P e Q (ambos são números primos grandes e diferentes), multiplicando por valores maiores que 400 por exemplo, depois calculamos o valor N com multiplicação de P por Q, em seguida calculamos o euler totiente:

$$Euler = (P-1) \times (Q-1) \quad \ldots (1)$$

Depois seleciona um inteiro E que seja co-primo do nosso euler totiente e procedemos a calculamos nosso D (Chave primária):

$$D^E \bmod Euler = 1 \quad \ldots (2)$$

Sabendo nosso D e E podemos criar as chaves públicas e privadas e fazer as encriptações de ida e volta conforme a necessidade:

Processo de ida:

Texto comum: T

$$\text{Texto cifrado: } C = T^E \bmod N \quad \ldots (3)$$

Processo de volta

Texto cifrado: C

$$\text{Texto decifrado: } M = C^D \bmod N \quad \ldots (4)$$

2.7 FAKER API

Faker API [8]é uma API de geração de dados utilizada para construções de projetos, testes unitários e criação de backends – cada dado gerado pelo fake é razoável em relação a um dado “real”. A API foi escrita primeiramente em PERL e depois portada para outras linguagens de programação, a API permite também que a formação desses dados seja baseada no país determinado pelo programador.

[7]Método acessado e estudado como base pode ser encontrado em: https://ieeexplore.ieee.org/document/4160265

[8]Faker API pode ser encontrado em: https://faker.readthedocs.io/en/master/

2.8 FUNÇÕES DE ATIVAÇÃO

Para criar um modelo de Machine Learning utilizando Neural Networks (NN) usa-se funções de ativação em suas entradas, saídas e principalmente nos *hiddens layers* (camadas escondidas), sendo essa a camada de computação e cálculos dos pesos, sendo responsável por pegar os dados de entradas, computá-los e ejetá-los para saída

Em cada camada existe um nodo contendo funções de ativações como outras configurações, como conexões, dimensões de entradas e total de saídas, os nodos de entrada tem como foco em providenciar dados para as camadas ocultas, por isso, os modelos desenvolvidos para apresentar neste artigo podem ter funções recorrentes entre eles, já as camadas escondidas possuem funções de ativação diferenciadas entre modelos, cada função realiza um cálculo diferente para avaliar os dados que passaram pelo pré-processamento e passar para a camada de saída onde as funções de ativação dessa camada são utilizadas para mostrarem os resultados da camada anterior, importante realçar as camadas escondidas são abstraídas e não são expostas de qualquer maneira.

3. OBJETIVOS

Neste artigo o objetivo principal proposto é uma abordagem visando garantir a proteção e segurança de dados privados seguindo a contextualização da L.G.P.D., transformando os incisos mencionados da lei em controladores de fluxo e utilizá-los em conjunto de aprendizagem de máquina, classificando dados alimentados ao algoritmo inteligente, visando proteger com encriptação os dados classificados como privados e rejeitar dados públicos, já que este não se encaixo dentro do contesto dos incisos do artigo 6 da lei apresentada.

Os objetivos secundários deste artigo são mostrar como as configurações dos modelos de máquinas supervisionados se comportam com configurações diferentes nos modelos e como essas configurações modificam a capacidade de identificar e realizar previsão dos dados privados e procurar uma aplicabilidade deste método no mundo real.

4. APLICAÇÕES

Uma das aplicações para esse modelo seria como um intermediador no database para encriptar os dados quando a detecção deles forem feitas, para evitar que eles possam ser vazados indevidamente. Considere uma situação onde invasores da rede de um sistema obterão acesso interno ao sistema, um administrador vigiando a rede ou até mesmo um sistema integrado de segurança poderia ativar o modelo para encriptar os dados sensíveis em uma chamada ao invés de fornecê-los desvalidos para os invasores, uma vez que dados encriptados não são fáceis de

quebrarem e dependendo do nível de encriptação, até mesmo impossíveis dentro de um determinado tempo, ou seja, em tempo de causar um dano real aos donos do sistema, já que estes tem tempo para ativar precauções, planos de contingência e avisar usuários do sistema sobre o comprometimento dos dados.

Outra possível aplicação é a utilização do modelo para a limitar acesso de dados para usuários de dashboards de dados tais como a Microsoft PowerBI, assim limitando o acesso das informações privadas por nível de usuário e tipo de requisição de dados, garantindo que apenas os operadores corretos de cada dado podem ter acesso aos dados.

5. METODOLOGIAS

A metodologia utilizada segue como – Realizar um processo de treino e desenvolvimento dos modelos, como é apresentado e demonstrado no fluxograma da Figura 1, primeiro é realizado o processo de geração de dados em uma base de dados vazia, pode também optar em escolher criar uma classe geradora de dados do qual no final do processo de geração de dados, este método foi o escolhido neste projeto. Esse processo de geração é feito utilizando o Faker API, este é uma API de geração de dados utilizada para construções de projetos, testes unitários e criação de base de dados – cada dado gerado pelo Faker é razoável em relação a um dado "real", isto é, ele possui as mesmas características que um dado privado real, mas não possui o vínculo do qual o dado privado real possui, está sendo a representação da privacidade de uma pessoa física, o dado gerado do Faker apresenta apenas a sua semelhança.

A API permite também que a formação desses dados seja baseada no país determinado pelo programador, permitindo vários tipos diferentes de dados gerados, como o foco é na LGPD, os dados gerados são de nacionalidade brasileira – Inicialmente gera-se os dados utilizando o Faker API, gerando dados considerados privados (supostamente relacionados diretamente com a privacidade pessoal de um usuário) e dados públicos (não relacionados diretamente com a privacidade do usuário). Após isso, os modelos criados e instanciados podem usar os dados criados pelo Faker API para serem treinados e testados.

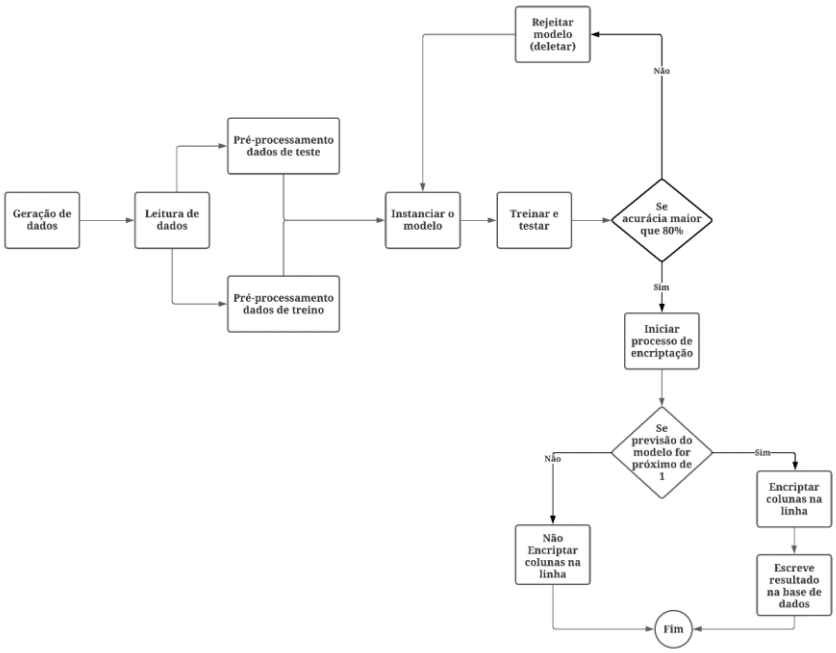
Antes de serem treinados e testados, os dados passam por um processo de pré-processamento para transformá-los em valores numéricos, para alimentar os modelos com informações tangíveis para eles, já que estes não conseguem compreender o alfabeto humano, essa transformação de dados é feita utilizando categorizações de cada coluna e depois pela transformação realizada por um Label Encoder, por fim os dados são separados em classes X (dados transformados) e Y (resultado dos dados), esse processo é feito para testes e treinos.

O teste e treino dos modelos são feitos depois de criar e compilar eles usando métricas de *binary crossentropy* para o *loss* e o otimizador adam, depois é feito o processo de treinamento e teste utilizando 225 épocas com um *batch size* de 25 utilizando call-backs com *earlystopping* monitorando os valores de perca, focando em minimizar as perdas e com uma paciência de 35 épocas, sempre que possível restaurar os melhores pesos.

Depois dos testes e treinamentos, o que determina um modelo ser escolhido para ir nas fases de previsão e encriptação são as métricas de acurácia, loss e métricas definidas como *True Posisitves* (Verdadeiros Positivos), *False Positives* (Falsos Positivos), *True Negatives* (Verdadeiros Negativos) e *False Negatives* (Falsos Negativos). Se a acurácia está acima de 80%, o modelo é considerado ótimo para as próximas etapas, porém é necessário garantir que o loss seja algo tangível e não muito alto, pois ser alto é uma indicação da configuração atual não ser a mais adequada para a base de dados geradas. As métricas apresentadas como os Verdadeiros Positivos serão discutidos na seção de desempenho dos modelos, após essa avaliação é iniciado a etapa de realização de previsão dos modelos.

Utilizando essa capacidade, se a previsão dos modelos for muito próxima de um, o dado é considerado privado, se muito perto de zero, o dado é considerado público. Dados privados são encriptados utilizando uma criptografia de RSA, permitindo o dado poder ser desencriptado, colaborando com o inciso VIII, os dados públicos são deixados como estão e por fim os dados são escritos na base de dados, linha a linha e mantendo a ordem de entrada da base de dados original.

Figura 1 – Fluxograma do SecureLearn.



Fonte: elaborado pelo autor.

6. ANÁLISES

O estudo realizado teve foco em verificar qual das configurações para os modelos podem obter o melhor desempenho em identificar dados privados e recomendar a sua encriptação, mostrando a capacidade dos modelos em serem ferramentas auxiliares para programadores encarregados de cuidar de bases de dados.

Os modelos foram feitos utilizando deep learning com redes neurais e funções de ativações variadas entre eles em busca da melhor configuração de funções e parâmetros para realizar a análise dos dados privados e públicos apresentados no modelo, com fim de realizar a encriptação das bases de dados geradas utilizando a capacidade de realizar previsões do modelo, sendo o melhor modelo, o eleito para ser uma das maneiras a serem utilizadas para aplicar sobre uma base de dados com características similares as montadas durante a experimentação do projeto.

O modelo 2 a ser apresentado na próxima seção deste artigo, foi o mais próximo de um desempenho considerado ideal para um *deploy* em um sistema com uma base de dados, devido a sua capacidade de identificar e recomendar encriptações baseado no seu vetor resultado de previsões, estas sendo probabilidades do modelo acreditar o dado estar na representação do qual ele pertence, sendo mais próximo de zero um dado considerado irrelevante ou público, e mais próximo de um, os dados considerados privados pela lei.

7. DESEMPENHO DOS MODELOS

Foram desenvolvidos um total de quatro modelos com funções de ativações variadas, cada modelo acessa os mesmos dados de teste e de treino para depois passarem pelo processo de avaliação de encriptação.

O primeiro modelo utiliza um tensor de 12 conexões com uma função de ativação ReLU e para as demais conexões a utilização do Hard Sigmóide para calcular os melhores pesos e realizar a saída das previsões, a tabela 1 possui a configuração do modelo.

Tabela 1 – Apresentação dos componentes do modelo 1 em forma e tabela.

Quantidade de Tensors	Tipo Ativação
12	ReLU
8	Hard Sigmoid
8	Hard Sigmoid
8	Hard Sigmoid
1	Hard Sigmoid

Fonte: elaborado pelo autor.

As funções de ativações deste modelo trabalham utilizando uma reta linear para diferenciar os dados de entrada, transportando-os para a camada de processamento do qual utiliza uma variante da função de Heaviside, podendo determinar em seu processamento saídas entre 0 e 1 após o processamento.

Após esse processamento/treinamento, o modelo é capaz de realizar previsões com cada saída em forma de porcentagem, indicando as chances daquela previsão está de acordo com o resultado real de cada dado, no caso, 0 ou 1 se não pertenceu ou pertence a categoria de dados sensíveis/privados respectivamente.

Tabela 2 – Comparação entre as detecções dos modelos apresentados no modelo 1, utilizando os parâmetros TP, FP, TN e FN.

Modelo 1: Hard Sigmoid.			
TP	**FP**	**TN**	**FN**
484	66	431	16

Fonte: elaborado pelo autor.

O modelo 2 utiliza funções de ativações baseadas na *Rectified Linear Activation Unit*, permitindo a ativação dos neurônios em tempo de processamento maior que outras quando os inputs não resultam em 0, diferentes das outras quais podem ter problemas para começarem a processar os dados na camada escondida , com isso os resultados desta versão apresentou resultados onde não existiam Falsos Negativos e apenas o parâmetros de Falsos Positivos acabaram sendo gerados, este podendo inferir no resultado de encriptação.

Sendo um Hard Sigmoide, essa camada pode não pegar todos os pontos de avaliação por ser uma função de ativação considerada não-suave, alguns dados podem terem sido evitados pela presença da avaliação binária.

Tabela 3 – Apresentação dos componentes do Modelo 2 em forma e tabela.

Quantidade de Tensors	**Tipo Ativação**
12	ReLU
12	ReLU
8	ReLU
4	ReLU
1	Hard Sigmoid

Fonte: elaborado pelo autor.

Ainda assim o modelo teve um resultado interessante quando comparado ao seu "irmão" anterior, principalmente olhando para os resultados na tabela 4. Porém a quantidade de Falsos Positivos detectados ainda foi relativamente alta olhando nessa mesma comparação.

Tabela 4 – Comparação entre as detecções dos modelos apresentados na versão final, utilizando os parâmetros TP, FP, TN e FN.

Modelo 2: Hard Rectified Linear Unit (ReLU).			
TP	**FP**	**TN**	**FN**
521	73	402	0

Fonte: elaborado pelo autor.

O modelo 3 é um modelo mais homogêneo usando apenas funções Sigmoide, sendo a mais famosa e permitindo uma flexibilidade maior, podendo ser utilizadas nas camadas de entrada, escondidas e nas camadas de saídas para parte de previsão. O sigmoid fornece também um gradiente mais suave evitando saltos quando retornando os resultados do processamento, utilizando saídas entre 0 e 1 sendo mais diretas com os resultados.

Tabela 5 – Apresentação dos componentes do Modelo 3 em forma e tabela.

Quantidade de Tensors	Tipo Ativação
12	Sigmoid
8	Sigmoid
8	Sigmoid
8	Sigmoid
1	Sigmoid

Fonte: elaborado pelo autor.

Tabela 6 – Comparação entre as detecções dos modelos apresentados na versão final, utilizando os parâmetros TP, FP, TN e FN.

Modelo 3: Pure Sigmoid			
TP	**FP**	**TN**	**FN**
517	4	399	75

Fonte: elaborado pelo autor.

Olhando apenas para a distribuição de detecções, este modelo obteve a melhor distribuição entre os modelos de forma geral, se aproximando do resultado desejado para o algoritmo de encriptação realizar a sua função, porém existem casos de detecções fora da curva que foram

categorizados errados. O modelo 4 é o mais heterogêneo de todos já apresentados, contendo três funções de ativação, ReLU para entradas, tangente para a camada escondida e uma *Hard Sigmoid* para a saída, essas escolhas foram feitas para entradas e saídas devido à como a Tangente se comporta na camada escondida, produzindo resultados entre [-1, 1], a ReLU vai pegar as entradas e trabalhar com elas num intervalo de [0, 1] e a saída usando *Hard Sigmoid* volta em a mostrar os resultados em [-1, 1], a tabela 7 mostrando a configuração do modelo.

Tabela 7 – Apresentação dos componentes do Modelo 4 em forma e tabela.

Quantidade de Tensors	Tipo Ativação
12	ReLU
8	Tangente
8	Tangente
4	Tangente
1	Hard Sigmoid

Fonte: elaborado pelo autor.

Tabela 8 – Comparação entre as detecções dos modelos apresentados na versão final, utilizando os parâmetros TP, FP, TN e FN.

Modelo 4: Tangente.			
TP	**FP**	**TN**	**FN**
521	83	393	0

Fonte: elaborado pelo autor.

Observando a tabela de detecções, mesmo alcançando o objetivo de não obter detecções falsas nos negativos, o modelo teve uma queda quando comparado ao modelo 2, onde obteve o menor resultado entre os dois quando olhando na coluna de Falsos Positivos podemos ver uma quantidade maior nas detecções do modelo 2.

Em todas as encriptações feitas pelos modelos utilizando o RSA customizado, as encriptações foram de acordo com as tabelas de detecções de cada modelo, incluindo os Falsos Positivos e Falsos Negativos, esse desenvolvimento será discutido na seção de limitações a seguir.

8. LIMITAÇÕES E DISCUSSÕES

No desenvolvimento do projeto não existiu muitas oportunidades de trabalhar com dados reais e tangíveis, as maiorias das oportunidades para manipula-los necessitaria de recurso monetário indisponível para o desenvolvimento desse projeto, então foi-se tomado a iniciativa

de utilizar a API chamada de Faker para desenvolver dados privados/sensíveis dos quais possam ser tangíveis o suficiente para desenvolver os modelos e treina-los em método de treinamento supervisionado para realizar o desejado e discutido no artigo, identificar o dado, encripta-lo se for sensível, deixa-lo visível se for irrelevante perante a lei.

Existe a possibilidade de os dados gerados apresentarem um bias para os modelos fazendo com que eles não consigam diferenciar adequadamente os tipos dos dados gerados, isso sendo um resultado da utilização da Faker API, como o método de pré-processamento (convertendo os valores baseados numa tabela dentro do Keras).

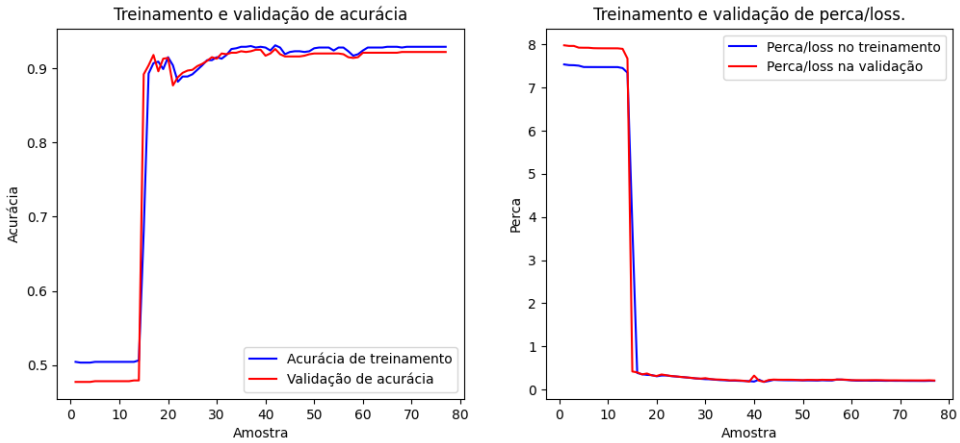
Outra das limitações desse projeto foi trabalhar com a API antiga do Tensorflow de processamento de texto, a nova API lançada no ano de 2023 apresenta uma capacidade de maior de processamento, interpretação e avaliação de texto – Dentro do contexto desse artigo, se a versão tivesse disponibilizada no tempo de desenvolvimento do projeto, os resultados dos modelos poderiam ser mais satisfatórios e apresentar variações mais interessantes.

O que pode ser dito também a respeito da capacidade de detecção de cada modelo quando observando as tabelas na seção de metodologia, é a sua acurácia em realizar uma previsão adequada, no caso pode ter ocorrido uma falta maior de coesão dos parâmetros das classes e no modelo para melhorar e possivelmente zerar essas detecções de Falsos Positivos e Falsos Negativos, olhando para o lado do tempo para desenvolver todos os modelos (em torno de 6 meses), os resultados apresentados foram interessantes e satisfatórios, apesar dos pontos apresentados.

Em desenvolvimento dessa área e desses tipos de modelos, é possível detectar a presença de picos de perdas/loss sendo onde o modelo pode ter dificuldade de validar os dados e acabam resultando em penalidades, perdas de melhores pesos e até avaliações erradas dos dados durante todo o processo.

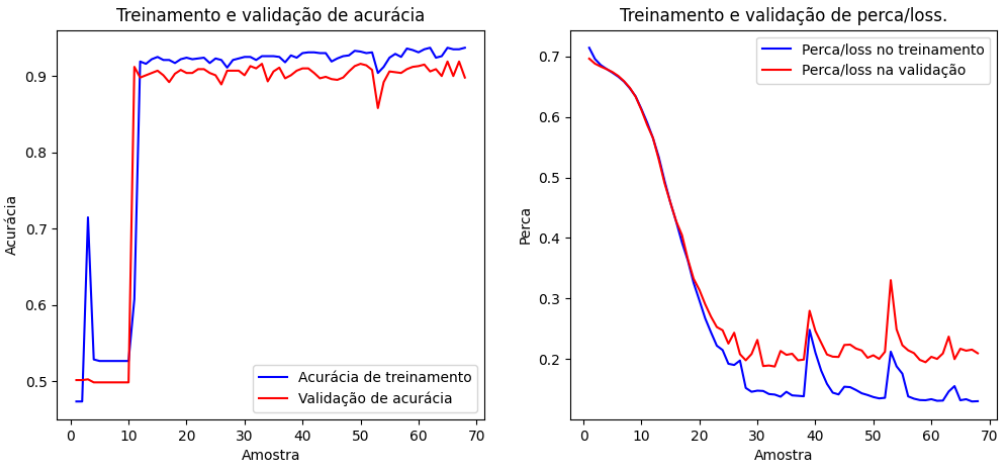
Os modelos 1, 3 e 4 foram vítimas dessas perdas, principalmente na parte de previsões, que é a parte mais relevante para a verificação das capacidades de cada um, em contrapartida o modelo 2 não obteve tais variações durante todo o seu processo de treinamento e validação, mas por não permitir análise de valores abaixo de zero durante o processamento na camada escondida, ocorreu detecções de FP e FN respectivamente, mesmo obtendo o melhor desempenho de distribuição de previsões e pouca perca durante sua execução.

Figura 2 – Desempenho gráfico do modelo 2.



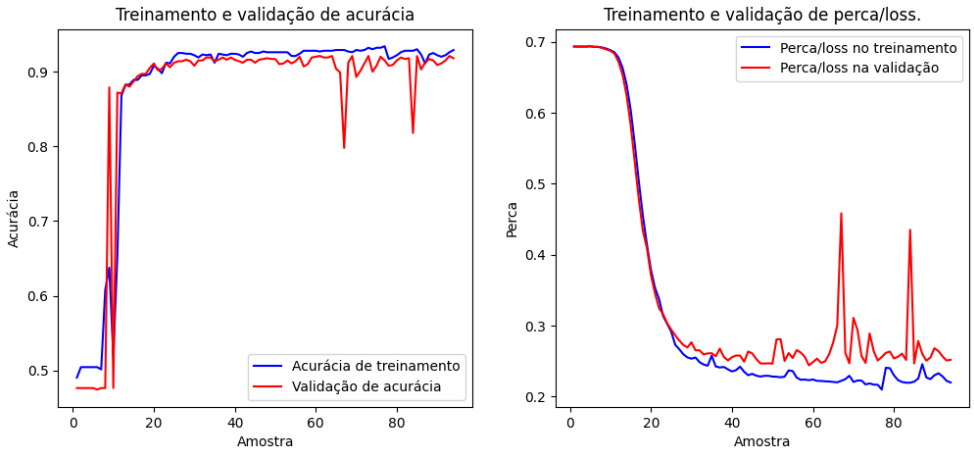
Fontes: Gráficos elaborados pelo autor.

Figura 3 – Desempenho gráfico do modelo 1.



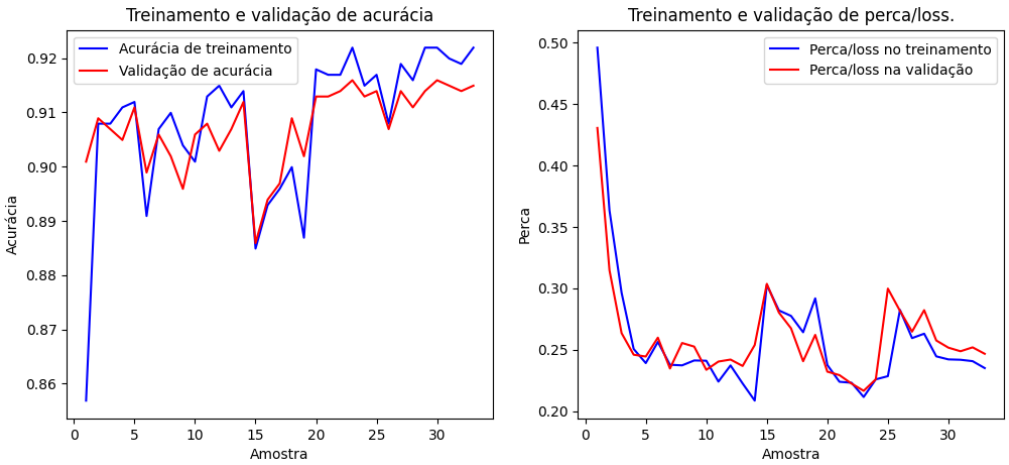
Fontes: Gráficos elaborados pelo autor.

Figura 4 – Desempenho gráfico do modelo 3.



Fontes: Gráficos elaborados pelo autor.

Figura 5 – Desempenho gráfico do modelo 4.



Fontes: Gráficos elaborados pelo autor.

Mesmo com o desempenho bem próximo do desejável o modelo 2 ainda teve problemas na detecção, voltando ao ponto de que os parâmetros não foram os suficientes para o processamento dos dados e para criação dos vetores de classes de X (classe de valores pré-processados) e Y (classe de resultados esperados), mostrando uma quantidade maior de parâmetros além das utilizadas.

Outro problema para o desenvolvimento do projeto foi a limitação física em hardware do computador utilizado para o desenvolvimento do projeto, já que muito dos modelos desenvolvidos disponíveis utilizam milhares e até milhões de parâmetros em conjunto de diversos processadores CUDA (*Compute Unified Device Architecture*) para mastigar e compreender os dados inseridos nos modelos.

9. TRABALHOS RELACIONADOS

A IBM realizou um trabalho onde foi desenvolvido um sistema automatizado na nuvem para detecção de irregularidades em uma base de dados, chamado de QRadar [2]. Este sistema realiza varreduras para confirmar se existem casos onde os dados sensíveis foram deletados com sucesso da base, se existir vestígios ou uma possibilidade de a deleção não ocorrer com sucesso, um trigger é ativado para realizar outra varredura mais profunda, os modelos apresentados anteriormente, realizar encriptação dos dados para estar de acordo com o inciso de permanência dos dados, já que caso necessário por investigação, eles precisam estar disponíveis, entretanto, os modelos não possui a capacidade de varreduras profundas como o modelo da IBM.

Neste trabalho de pesquisa [3], também da IBM, o modelo de ML utilizado por pelos pesquisadores usa escolhas binárias para os dados coletados serem minimizados de acordo com

a GDPR, no caso, estão a utilizar dado de um hospital para escolher quais dados podem ser deletados baseado nos parâmetros escolhidos. Comparado aos modelos, apesar de um resultado mais condizente com a LGPD, existe uma porcentagem pequena de valores não encriptados como falsos negativos, onde neste projeto, pelo escopo mais simples, foi possível atingir o corte dos dados necessários dos salários para aparecer na dashboard de dados.

10. TRABALHOS FUTUROS

Caso um possível desenvolvedor se interesse em contribuir para essa ideia apresentada neste artigo, recomenda-se procurar aprimoramentos dos parâmetros de análise, de entrada em cada layer de entrada e utilização, se possível, de ambientes de processamento conectados na nuvem, pois este projeto foi realizado localmente utilizando um computador relativamente capacitado para desenvolvimento e teste dos modelos, esta limitação não ocorre em ambientes computacionais na nuvem, pois estes funcionam baseados no tempo de uso de cada modelo e o tempo de processamento dos dados.

Outra melhoria é a utilização das novas ferramentas, API's e frameworks desse projeto, nos seis meses de desenvolvimento até a escrita do artigo, recentemente houve uma melhoria nas capacidades de processamento de dados textuais e de valores pelo Tensorflow, devido a apresentação de competidores tais como o ChatGPT.

Mencionando ChatGPT, poderia ser feito uma API's intermediária entre a IA e os dados, já que o ChatGPT possui uma quantidade de parâmetros aproximado em milhões, tendo uma capacidade maior de entender e processar os tipos de dados pessoais. Um adendo e aviso a este método de desenvolvimento, o ChatGPT não possui uma categoria de privacidade nos seus pesos, ou seja, pode existir a chance dos dados sensíveis se manterem dentro do modelo até após sua execução e desligamento, resultando em um evento do qual um outro usuário qualquer pode fazer uma requisição aleatória e o ChatGPT acabar entregando os dados sensíveis como foram os casos de programadores pedindo ajudas com códigos e a IA entregando códigos protegidos por *Defense Contract Management Agency* (DCMA), não só desrespeitando a lei, como entregando códigos protegidos por Propriedade Intelectual (IP), caso o responsável pelo ChatGPT resolvam esse problema, a utilização dessa IA é interessante.

11. CONCLUSÕES

Neste artigo foi apresentado como a internet integrada a serviços para usuários se tornou mais comum e cada vez mais presente na vida de maneira geral e as consequências desses sistemas não utilizarem mecanismos (seja de direito ou da parte da computação) até a criação da

LGPD, Machine Learning e como a área de Ciências da Computação junto com Direito se intersectam para garantir a integridade dos dados sensíveis de usuários em diversos sistemas integrados na internet.

O Faker API foi importante no desenvolvimento desse projeto não só pela geração de dados utilizados para o desenvolvimento dos modelos, como na influência individual em cada modelo para a detecção dos dados.

Os modelos apresentados tiveram a capacidade de computar e compreender os dados para gerar resultados encriptados de dados pessoais e deixar dados irrelevantes em abertos, pois estes não são categorizados pela lei como dados sensíveis, por não haver uma ligação pessoal entre a informação presente na database e a uma pessoa, onde o dado diz algo sobre a vida íntima da pessoa em questão. Com isso pode se afirmar que existe sim um espaço do qual a área de IA pode preencher para ajudar profissionais de direito e profissionais de computação a realizarem seus trabalhos com facilidade e automação, como também garantir a integridade da lei, assim como expandir as áreas de oportunidades no mercado de trabalho pela existência dessa nova intersecção entre as áreas apresentadas.

Os modelos construídos, testados e estudados nesse artigo são exemplos de como a área de IA pode ajudar nessa aplicação da lei, como facilitar o trabalho judicial na detecção de dados não respeitando a integridade da lei, como a aplicação de incisos da lei para garantir a proteção dos dados e protegê-los com encriptação.

REFERÊNCIAS BIBLIOGRÁFICAS

Boutaba, Raouf; Mohammad A., Salahuddin; Limam, Noura; Ayoubi, Sara; Shahriar, Nashid; Estrada-Solano Felipe; Caicedo M. Oscar; **"A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities"**, p. 1-99, 2018. Disponível em: <https://jisajournal.springeropen.com/articles/10.1186/s13174-018-0087-2#Sec2>

Cots Marcio, Oliveira Ricardo, **"Lei Geral de Proteção de Dados Pessoais Comentada"**, Jusbrasil, Disponível em: <https://www.jusbrasil.com.br/topicos/200399469/artigo-6-da-lei-n-13709-de-14-de-agosto-de-2018>, Acessado em: 08 de abril de 2024.

Fatih, Ertam; Galip, Aydin; **"Data classification with deep learning using TensorFlow"**, p. 1-4, 2017. Disponível em: <https://ieeexplore.ieee.org/document/8093521>

Filipe Lima Rapôso, Cláudio; Melo de Lima, Haniel; Ferreira de Oliveira Junior, Waldecy; Aragão Ferreira Silva, Paola; Elaine de Souza Barros, Elaine; **"LGPD - LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS EM TECNOLOGIA DA INFORMAÇÃO: Revisão Sistemática"**, p. 1-10. 2019. Disponível em:<https://revistas.cesmac.edu.br/index.php/administracao/article/view/1035>

Fortunato, Caroline, **"Using QRadar for LGDP"**, 19, julho de 2019, Disponível em:<https://www.proof.com.br/wp-content/uploads/2019/08/Using-QRadar-for-LGPD.pdf>

GDPR, **"General Data Protection Regulation GDPR":** OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018. Disponível em: <https://gdpr-info.eu/> Acesso em: 27, agosto de 2022

Goldsteen, Abigail; Ezov Gilad; Shmelkin, Ron; Moffie, Micha; Farkash, Ariel; **"Data minimization for GDPR compliance in machine learning models"**, p. 1-15, 2021, Disponível em: <https://link.springer.com/article/10.1007/s43681-021-00095-8>

International Conference on Information and Communications Technology (ICOIACT), **"A Comparative Study MD5 and SHA1 Algorithms to Encrypt REST API Authentication on Mobile-based Application"**, 2019ª, p. 206-211, doi: 10.1109/ICOIACT46704.2019.8938570.

J. D. Hunter, **"Matplotlib: A 2D Graphics Environment"**, in Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, May-June 2007, doi: 10.1109/MCSE.2007.55. Disponível em:<https://ieeexplore.ieee.org/document/4160265>

Leal da Silva, Julia; **“Tomada de Decisão Automatizada e Controle pela LGPD”.** IAPD, 20, janeiro de 2021. Disponível em: <https://iapd.org.br/decisao-automatizada-lgpd-direito-aexplicacao/> Acesso em: 24, agosto de 2022.

Leal da Silva, Julia; **“Tomada de Decisão Automatizada e Controle pela LGPD.** IAPD, 20, janeiro de 2021. Disponível em: <https://iapd.org.br/decisao-automatizada-lgpd-direito-aexplicacao/> Acesso em: 24, agosto de 2022.

McCallum**,** Shiona; **“ChatGPT Banned in Italy over privacy concerns”**, Site: Notícia, abril de 2023, Disponível em: <https://www.bbc.com/news/technology-65139406> Acesso em: 16/06/2023.

Nguyen Quang-Hung; Hieu Doan; Nam Thoai; **“Performance Evaluation of Distributed Training in TensorFlow 2”**, p. 1-5, 2020. Disponível em: < https://ieeexplore.ieee.org/document/9353085 >

NUMPY PROJECT AND COMMUNITY, **“What is NumPy?”**, 2022ª.

PANDAS, **“10 minutes to pandas”,** Site: User Guide, 2022. Disponível em: <https://pandas.pydata.org/docs/user_guide/10min.html#viewing-dataAcesso em: 24 de agosto de 2022.

PANDAS, **“Pandas DataFrame”** Site: About, 2022. Página sobre nós. Disponível em: <https://pandas.pydata.org/about/> Acesso em: 24 de agosto de 2022.

Pavani and P. Sriramya, **“Enhancing Public Key Cryptography using RSA, RSA-CRT and N-Prime RSA with Multiple Keys”**, 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1-6, doi: 10.1109/ICICV50876.2021.9388621. Disponível em: <https://ieeexplore.ieee.org/document/9388621>

PYTHON SOFTWARE FOUNDATION. **“Python Language”,** Site: Documentation, 2022. Disponível em: <https://www.python.org/doc/> Acesso em: 24 de agosto de 2022

Roberto Fernandes Castilho, José; Andrade Gomes, Henrique; **“Legislação Básica de Direito da Informática”**, 2° edição reformulada e atualizada, São Paulo, Editora Pillares.

Shanmugam, Divya; Shabanian, Samira; Diaz, Fernando; Finck, Michèle, Biega, Asia; **“Learning to Limit Data Collectionvia Scaling laws: A computional Interpolation for the Legal Principle of Data Minimization”**, p. 1-11, 2022. Disponível em: < https://arxiv.org/abs/2107.08096 >

Song, Congzheng; Ristenpart, Thomas; Shmatikov, Vitaly; "**Machine Learning Models that Remember Too Much"**, p. 1-15, 22, setembro de 2017. Disponível em: <https://arxiv.org/pdf/1709.07886.pdf>

Spadaccini de Teffé, Chiara; Viola, Mario; **"Tratamento de dados pessoais na LGPD: estudo sobre as bases legais"**, p. 1-38, 2020. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/510>

Stallings, William; Brown, Lawrie; **"Computer Security: Principles and Practice Second Edition"**, 2° edição reformulada e atualizada, São Paulo, Pearson; 2ª edição.

Tankard, Colin; **"What the GDPR means for business"**, p. 1-8, 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1353485816300563?casa_token=pjOHlq5iYsAAAAA:dhlGukGx-SanwjnBY9aPtw36O1CslOJG1wZ7wTEryTJHi5QtFVM2G6kc8CwuemdVrUDRl2tRqJoA>

Vieira Souza, **"Aplicações de software desenvolvidas no contexto da Inteligência Artificial (IA), Machine Learning e Big Data e o direito dos cidadãos de acordo com a Lei Geral de Proteção de Dados (LGPD)"**, 2021ª, p. 1-83. Disponível em: <https://bdm.unb.br/bitstream/10483/30275/1/2021_IuriSousaVieira_tcc.pdf>