Aprendizado de máquina sobre dados médicos com preservação de privacidade

Felipe Alves Barusso[1], **Bruno Bogaz Zarpelão**[1]

[1]Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

felipe.barusso@uel.br, brunozarpelao@uel.br

Abstract. *In today's medical services, an individual's data is often stored digitally. With this data, it is possible to apply machine learning algorithms to automate processes and perform analysis on them. However, these operations usually have a high computational cost. Thus, health institutions seek to outsource the data processing. However, due to the sensitive nature of the information contained therein, it is necessary to guarantee the privacy of the data when moved to an external server. In this work, a way to move data to multiple external processing environments is proposed, so that servers do not have contact with the original information in its plain form.*

Resumo. *Nos serviços médicos atuais, os dados de um indivíduo são muitas vezes armazenados de forma digital. Com estas informações, é possível aplicar algoritmos de aprendizado de máquina para automatizar processos e realizar análises sobre elas. Porém, estas operações costumam possuir um alto custo computacional. Assim, as instituições de saúde procuram terceirizar o processamento destes dados. Contudo, devido a natureza sensível da informação neles contida, é necessário garantir a privacidade dos dados ao serem enviados a um servidor externo. Neste trabalho, é proposta uma maneira de transportar os dados para múltiplos ambientes externos de processamento, de maneira que os servidores não tenham contato com a informação original.*

1. Introdução

A transição dos serviços de saúde para o meio digital apresentou benefícios como processos mais eficientes e a diminuição de custos, levando ao surgimento de novos serviços e modelos de negócios [9]. Informações sobre consultas, dados de saúde e imagens de exames de um paciente são armazenadas nos bancos de dados de instituições médicas.

Com estes dados, é possível aplicar técnicas de aprendizado de máquina para atingir resultados relevantes. Uma possível aplicação é um algoritmo que automatiza parte de um diagnóstico. Outro exemplo, é um modelo de classificação de imagens que identifica tecidos malignos em imagens de núcleos de células mamárias [3].

Porém, com a alta quantidade de dados presentes em uma instituição de saúde, o custo computacional de aplicar um algoritmo de aprendizado de máquina sobre estes dados é alto. Portanto, as instituições procuram realizar o processamento das informações em servidores externos. Este fato faz surgir riscos de privacidade devido à sensibilidade da informação contida em dados de saúde.

Dados médicos frequentemente contém informações que podem revelar a identidade de um paciente, o que torna o controle de acesso aos mesmos um fator importante. Além disso, os dados de saúde pessoal têm um alto valor para as empresas farmacêuticas, seguradoras e empregadores, o que os tornam alvo de ataques [4].

Em agosto de 2021, a empresa americana de segurança em nuvem *Wiz* encontrou uma vulnerabilidade crítica na plataforma *Microsoft Azure*, que permitia o controle total e remoto de contas de outros usuários do banco de dados da Azure, o *Cosmos DB*. Alguns clientes do *Cosmos DB* incluem: *Coca-Cola*, *ExxonMobil* e *Walgreens*. Este fato mostra que nem mesmo empresas com um alto investimento em segurança estão imunes a vazamento de dados [6].

Uma possível solução para a preservar a privacidade da informação ao enviá-la para servidores externos é o conceito de compartilhamento de segredo. A instituição que deseja fazer o uso de ambientes externos de processamento primeiro divide os dados e acrescenta um ruído em cada parte. Assim, cada servidor de processamento recebe dados que não revelam informações sobre o conjunto original. Por fim, os servidores devolvem os resultados que devem ser reunidos e decodificados para revelar a informação desejada.

Neste projeto, será implementada uma solução de preservação de privacidade baseada no compartilhamento de segredo, para que o processamento de dados médicos possa ser distribuído em servidores externos sem que haja risco de segurança.

Por fim, será verificado se existe diferença significativa em termos de tempo de execução e acurácia entre um algoritmo de aprendizado de máquina que faz o uso da técnica de preservação de privacidade e um mesmo algoritmo que não utiliza a técnica.

Este documento contém as seguintes seções:

- A seção 2 apresenta os conceitos relevantes ao trabalho, assim como sua relação ao desenvolvimento.
- A seção 3 apresenta as metas que o trabalho deseja atingir.
- A seção 4 apresenta como se pretende atingir as metas da seção 3 em detalhe.
- A seção 5 apresenta o cronograma de execução.
- A seção 6 apresenta o que se deseja concluir com este trabalho.

2. Fundamentação Teórico-Metodológica e Estado da Arte

2.1. Aprendizado de máquina

O aprendizado de máquina é o estudo científico de algoritmos e modelos estatísticos que realizam uma tarefa específica sem serem explicitamente programados para tal fim [5]. Este conceito pode ser aplicado para múltiplos propósitos como mineração de dados, processamento de imagens, analítica preditiva etc.

Dado um conjunto de dados, nem sempre é possível extrair informações relevantes de forma trivial. Para isso, é possível empregar algoritmos de aprendizado de máquina para inferir relacionamentos entre os dados. É importante ressaltar que existem diversas abordagens ao aprendizado de máquina, sendo que o emprego de cada uma é situacional e depende do contexto em que é utilizada.

A aprendizagem supervisionada, na qual este projeto se foca, consiste em induzir um modelo a partir de um conjunto de entradas e saídas, sendo dividida em dois tipos de algoritmos:

1. Classificação: é utilizado para classificar observações em categorias específicas. Ele reconhece entidades específicas dentro do conjunto de dados e tenta tirar conclusões sobre como essas entidades devem ser rotuladas.
2. Regressão: é utilizado para entender a relação entre variáveis dependentes e independentes.

Na etapa de treinamento, o algoritmo recebe pares de entradas e saídas com valores conhecidos. Ele utiliza os dados para criar uma função que mapeie cada entrada com sua saída respectiva. Na medida que o algoritmo recebe pares de entrada/saída, ele atualiza os pesos da função para criar uma relação com mais acurácia. Esta atualização de modelo é realizada até que um parâmetro de convergência seja atingido.

Na etapa de predição, que ocorre após o treinamento do modelo, o algoritmo passa a receber apenas os dados de entrada, retornando uma saída estimada. Ao fornecer entradas com saídas conhecidas, é possível deduzir a taxa de acurácia do modelo através da seguinte fórmula:

$$Acurácia = \frac{Total\ de\ classificações\ corretas}{Total\ de\ entradas}$$

2.2. Compartilhamento de segredo

O compartilhamento de segredo é um método de criptografia utilizado na troca de informações. Um segredo é dividido entre um conjunto de participantes, onde cada um recebe apenas uma parte do segredo. Em seguida, na fase denominada recuperação ou reconstrução, um subconjunto de um número predefinido de participantes colabora para revelar o segredo. Na Figura 1, mostramos uma visão geral do processo de compartilhamento de segredos.

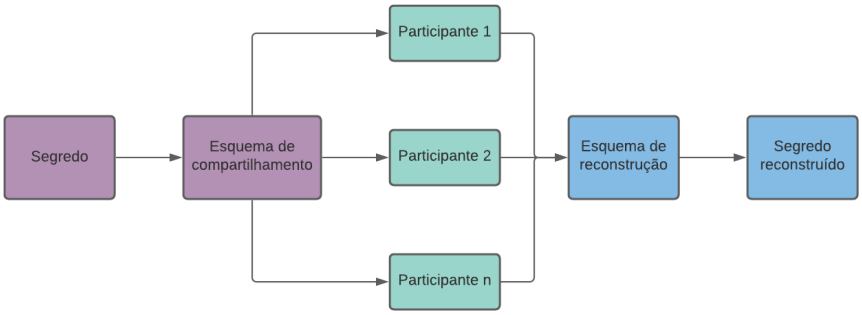

Figura 1. Processo de compartilhamento e reconstrução de um segredo

Formalmente, a informação inicial (denominada de segredo) é dividida em um conjunto de n participantes. Cada participante pode receber uma parte genuína de informação ou uma ”sombra”(ruído). O objetivo é que não seja possível realizar inferências sobre a informação original a partir das subdivisões, o que garante que ninguém terá o monopólio do segredo. Um subconjunto predefinido que consiste em t participantes (onde $t < n$) pode colaborar para revelar de volta a informação criptografada [7].

Este trabalho tem como foco o método proposto por Adi Shamir [8] em 1979: um esquema de compartilhamento de segredo chamado *Polynomial Secret Sharing Scheme*,

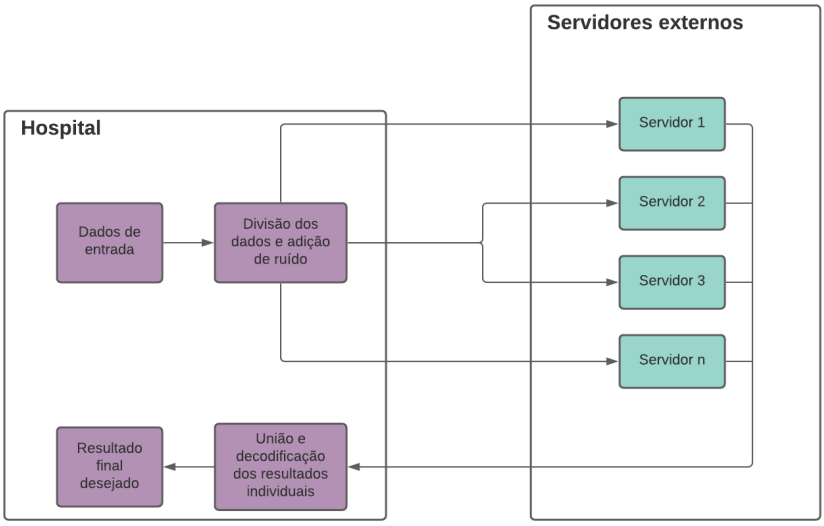

Figura 2. Esquema de compartilhamento de segredo contextualizado em um cenário onde um hospital terceiriza o processamento de dados médicos

no qual o autor demonstra um método para compartilhar um inteiro secreto a_0 utilizando um polinômio de grau *t - 1*.

É possível contextualizar o compartilhamento de segredo em um cenário de saúde. Uma instituição pode querer terceirizar o processamento de um conjunto de dados médicos em múltiplos servidores externos. Para isso, ela pode optar por utilizar o compartilhamento de segredo para que cada um desses servidores externos não tenha acesso aos dados originais, mas sim a uma subdivisão com ruído adicionado dos mesmos. Na Figura 2, mostramos como um hospital empregaria o conceito de compartilhamento de segredo.

2.3. Privacidade de dados médicos

Com o advento da era de serviços de saúde digitais, é importante utilizar dados de pacientes para realizar análises em profundidade e tratamentos personalizados. No entanto, os limites da infraestrutura das organizações de saúde e as ameaças de vazamento de dados colocam obstáculos no compartilhamento de dados médicos.

Por motivos de segurança, as instituições tem optado por construir seus sistemas de saúde em um domínio fechado com medidas defensivas, como uma rede privada equipada com *firewalls* e sistemas de detecção de intrusão [2].

Porém, a dimensão destes conjuntos de dados aliada ao custo computacional de realizar analises complexas, faz com que as instituições procurem utilizar serviços externos para realizar o processamento de informações. Assim, um método de garantir a privacidade de dados médicos ao enviá-los para servidores externos é de alta importância.

3. Objetivos

Este trabalho tem como objetivo geral implementar um método de aprendizado de máquina com preservação de privacidade tendo como base um conjunto de dados

médicos. Além disso, o projeto procura também avaliar como a preservação de privacidade pode impactar na acurácia e no tempo de processamento do algoritmo de aprendizado. Mais especificamente, o projeto busca:

1. Fazer um levantamento da literatura a respeito da área de privacidade de dados de saúde e de técnicas de aprendizado de máquina com preservação de privacidade.
2. Selecionar um conjunto de dados de caráter médico que seja compatível com o método a ser implementado.
3. Implementar um algoritmo de aprendizado de máquina supervisionado sem preservação de privacidade como *baseline*.
4. Implementar um método de aprendizado de máquina supervisionado com preservação de privacidade baseado em compartilhamento de segredos.
5. Realizar a análise dos resultados.

4. Procedimentos metodológicos/Métodos e técnicas

Inicialmente será realizado o levantamento bibliográfico sobre os temas a serem abordados no trabalho. Isso inclui o campo de privacidade de dados de modo geral, proteção de dados médicos em específico e algoritmos de classificação que possuem a característica da preservação de dados.

Em seguida, é necessário escolher uma base de dados para treinar o modelo de aprendizado de máquina. É importante que o conjunto de dados selecionado seja compatível com o algoritmo de classificação a ser implementado. Além disso, os dados devem ter uma característica sensível (característica intrínseca dos dados médicos), para que faça sentido a preservação dos mesmos.

Por este motivo, foram selecionados conjuntos de dados do *Breast Cancer Wisconsin* [10]. Os dados são extraídos a partir de uma imagem digitalizada de uma operação chamada *fine needle aspirate* de massa mamária. Eles contém informações a respeito do núcleo das células.

Assim, com a base de dados selecionada, é necessário realizar a implementação de um algorítimo de classificação com regressão linear para criar uma base comparativa. Então, será realizada a implementação de um método de aprendizado de máquina supervisionado com preservação de privacidade baseado em compartilhamento de segredos. Ambos algorítimos serão desenvolvidos em *Python*.

A segunda implementação tem por objetivo realizar o treinamento de um modelo de classificação utilizando o conceito de compartilhamento de segredo. A etapa de treinamento será realizada através da divisão do conjunto de dados original. Por conseguinte, cada repartição será enviada para uma simulação de servidores distribuídos.

Cada servidor realizará os cálculos necessários sobre os dados que lhe foram entregues. É importante ressaltar que com a informação que cada servidor recebe, não é possível inferir sobre o conjunto de dados original. Por fim, cada servidor devolve seus resultados para o ponto original, que reuni os valores e faz as operações necessárias para chegar na resposta desejada.

Por fim, será realizada uma análise sobre os resultados obtidos. Será possível comparar os resultados da implementação que emprega a técnica de preservação de privacidade com o algorítimo *baseline* em fatores como acurácia e tempo de execução.

5. Cronograma de Execução

Atividades a serem realizadas:

1. Levantamento bibliográfico;
2. Estudo aprofundado da biblioteca *TF Encrypted*[1];
3. Seleção do conjunto de dados;
4. Implementação do algoritmo sem preservação de privacidade (*baseline*);
5. Implementação do algoritmo com preservação de privacidade;
6. Testes no conjunto de dados;
7. Análise dos resultados;
8. Escrita do TCC;

Tabela 1. Cronograma de Execução

	set	out	nov	dez	jan	fev	mar	abr
Atividade 1	x	x						
Atividade 2		x						
Atividade 3			x					
Atividade 4			x	x				
Atividade 5			x	x	x			
Atividade 6					x	x		
Atividade 7						x	x	
Atividade 8					x	x	x	x

6. Contribuições e/ou Resultados esperados

Com este trabalho, espera-se verificar se existe diferença significativa entre um algoritmo de aprendizado de máquina que faz o uso de uma técnica de preservação de privacidade e o mesmo algoritmo que não utiliza a técnica, em termos de tempo de execução e acurácia para um conjunto de dados médicos.

7. Espaço para assinaturas

Londrina, *12 de setembro de 2022*.

Felipe Alves Barusso

Aluno

Orientador

Referências

[1] Encrypted deep learning in tensorflow. `https://tf-encrypted.io/`. Acessado em: 2022-09-11.

[2] Hao Jin, Yan Luo, Peilong Li, and Jomol Mathew. A review of secure and privacy-preserving medical data sharing. *IEEE Access*, 7:61656–61669, 2019.

[3] Laila Khairunnahar, Mohammad Abdul Hasib, Razib Hasan Bin Rezanur, Mohammad Rakibul Islam, and Md Kamal Hosain. Classification of malignant and benign tissue with logistic regression. *Informatics in Medicine Unlocked*, 16:100189, 2019.

[4] Jingquan Li. Ensuring privacy in a personal health record system. *Computer*, 48(2):24–31, 2015.

[5] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386, 2020.

[6] Joseph Menn. Microsoft warns thousands of cloud customers of exposed databases. *Reuters*, 2021.

[7] Parsa Sarosh, Shabir A Parah, and Ghulam Mohiuddin Bhat. Utilization of secret sharing technology for secure communication: a state-of-the-art review. *Multimedia Tools and Applications*, 80(1):517–541, 2021.

[8] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[9] Xiang Su, Jarkko Hyysalo, Mika Rautiainen, Jukka Riekki, Jaakko Sauvola, Altti Ilari Maarala, Harri Hirvonsalo, Pingjiang Li, and Harri Honko. Privacy as a service: Protecting the individual in healthcare data processing. *Computer*, 49(11):49–59, 2016.

[10] W H Wolberg and O L Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences*, 87(23):9193–9196, 1990.