

Desafios no compartilhamento de dados de saúde entre entidade de saúde privada e a academia para fomentar a pesquisa e produção científica

Emanuela C. R. Gonçalves¹, Guilherme D. Juraszek²
Marco Antunes², Mateus L. Bittencourt^{2,3}

¹Hospital de Olhos Sadalla Amin Ghanem (HOSAG)
R. Camboriú, 35 – Glória – Joinville, SC.

²Magrathea Labs
R. Orestes Guimarães, 876 – América – Joinville, SC.

³Universidade do Estado de Santa Catarina (UDESC)
R. Paulo Malschitzki, 200 – Zona Industrial Norte – Joinville, SC.

{guilherme,mateus.bittencourt}@magrathealabs.com

marco@antunes.consulting, emanuela.goncalves@sadalla.com.br

Abstract. *O crescimento e a velocidade de produção de dados - que podem ser extremamente úteis para pesquisa científica - nas empresas de diversos ramos tem sido bem divulgada nos últimos tempos (a chamada era do Big Data), porém, o que passa despercebido normalmente é a dificuldade da 'academia' em ter acesso a esses dados de forma a permitir a sua utilização em pesquisa dado o imenso abismo que separa empresas e universidades. Nesse contexto, irei relatar minha experiência no projeto de disponibilização de dados de saúde (exames oftalmológicos de Glaucoma) do Hospital de Olhos Sadalla Amin Ghanem (HOSAG) para alunos da Universidade do Estado de Santa Catarina (UDESC) utilizarem em pesquisa.*

Audiência: gerentes de projeto, consultores, cientistas de dados, professores, pesquisadores.

Breve biografia dos autores:

- Marco Antunes (Apresentador do Trabalho): Tecnólogo em Processamento de Dados (UnP), Pós-graduado em Gestão Empresarial (FGV), pós-graduado em Gerenciamento de Projetos (FGV), Data Scientist (Datacamp). 22 anos de experiência profissional desenvolvendo software de ERP Datasul/TOTVS para grandes empresas, 3 anos desenvolvendo soluções de inteligência artificial/analytics como consultor de mercado. LinkedIn: <https://www.linkedin.com/in/antunes73/>;
- Mateus Leite Bittencourt: Acadêmico do curso de Bacharelado em Ciência da Computação (UDESC), Trainee de Engenharia de Dados na Magrathea Labs. Trabalho de Conclusão de Curso em andamento no desenvolvimento de uma Rede Neural Artificial para classificação de imagens de Glaucoma, da base de dados do HOSAG. LinkedIn: <https://www.linkedin.com/in/mateus-leite-bittencourt-33956216b/>.

1. Introdução

A cada dia são gerados uma grande quantidade de dados por diversas empresas e segmentos da indústria, muitas vezes dados importantes para contribuição científica se explorados de maneira apropriada por instituições de ensino e pesquisa. Porém muitas vezes existe uma dificuldade para que universidades e outras instituições de pesquisa possam ter acesso aos dados de empresas e utilizá-los.

Este relato visa descrever alguns desafios enfrentados ao realizar a extração e disponibilização de dados do Hospital de Olhos Sadalla Amin Ghanem (HOSAG) para projetos de pesquisa da Universidade do Estado de Santa Catarina (UDESC). Os dados em questão foram um conjunto de imagens do fundo de olho mais algumas informações sobre o perfil do paciente para projeto de pesquisa envolvendo a classificação da presença ou não da doença de Glaucoma nas imagens, por meio do uso de *Deep Learning*.

2. Glaucoma

Entre as doenças que mais causam deficiência visual em estágios mais avançados até a cegueira, está o Glaucoma, uma neuropatia óptica causada principalmente pelo aumento da pressão intraocular (PIO), resultando em uma perda gradativa do campo visual, até na cegueira total nos casos mais graves [Mantravadi and Vadhar 2015].

Apesar de não haver cura para o Glaucoma, existe tratamento para esta doença, por isso a importância de um diagnóstico precoce, o que potencializa as chances de controle e estabilização da doença por meio de tratamento e acompanhamento por profissionais [Conlon et al. 2017]. Uma das formas de diagnóstico do Glaucoma é através da análise de uma retinografia / imagem médica de Fundo de Olho (FO) [Dorion 2002]. A Figura 1 ilustra essas imagens médicas de um olho saudável e com Glaucoma, respectivamente.

Com acesso às imagens (em grande volume) de exames de Fundo do Olho, pesquisadores poderão utilizar técnicas de *Machine Learning (Deep Learning)* para análise de padrões das imagens no estudo de diagnósticos automatizados mais precisos e rápidos.



(a) Olho Saudável



(b) Olho com Glaucoma

Figura 1. Exames de Fundo de Olho (HOSAG) de um olho sem e com a doença de Glaucoma, respectivamente.

3. Desafios Encontrados

3.1. Confiança

Inicialmente foi necessário que houvesse um reconhecimento mútuo de percepção de valor entre as instituições envolvidas (HOSAG como disponibilizadora dos dados em quantidade e qualidade necessários e UDESC como instituição de ensino) e um alinhamento de propósitos para que pudesse ser desenvolvido um projeto de colaboração.

- O HOSAG procurando explorar o potencial dos dados que acumulou ao longo dos anos com sua operação de alto nível para a pesquisa de novas formas de diagnóstico através dos dados;
- A UDESC oportunizando para seus alunos o acesso a dados em volume e de qualidade para trabalhos de pesquisa;
- Nesse ponto destaco o papel da Magrathea Labs (prestadora de serviços de consultoria de dados e software) como entidade catalisadora do projeto que operacionalizou as atividades necessárias através seu conhecimento técnico.

3.2. Entendimento das Informações

Uma vez alinhada a intenção entre as entidades, era hora de identificar e selecionar quais informações seriam utilizadas no projeto. Mas o conjunto de dados de exames que o HOSAG possui é muito grande e diverso com muitas aplicações possíveis. Então foi necessário um repasse, pelo time do HOSAG, para o time da UDESC e da Magrathea Labs para que houvesse um entendimento básico (sendo as informações sobre os exames e doenças de significativa complexidade em função de sua especialização) e que permitisse a seleção das informações mais adequadas ao projeto. O resultado dessa etapa foi a seleção da doença de Glaucoma como aplicação mais interessante em função dos dados disponíveis que eram:

- Imagens de exames de fundo de olho;
- Os resultados desses exames (classificação em diagnóstico positivo ou negativo para a doença);
- Dados básicos do perfil dos pacientes (idade e sexo).

Partimos então para a identificação dos dados disponíveis, em quais formatos estavam armazenados (imagens, dados tabulares e dados descritivos), e como se relacionam, por exemplo, como identificar o resultado do diagnóstico de uma determinada imagem - pois estavam em sistemas diferentes, assim também como entender a codificação de nomenclatura dos arquivos.

3.3. Entendimento dos Sistemas Envolvidos

Nesta etapa identificamos quais sistemas utilizados no HOSAG estariam envolvidos. Aqui ficou evidente que esses sistemas clínicos são muito especialistas e que não foram desenvolvidos com vista ao compartilhamento de dados como citaremos a seguir.

Sistema que registra as imagens dos exames de fundo de olho. Neste caso até havia uma funcionalidade de exportação de dados porém de forma individual, ou seja, precisávamos selecionar um paciente de cada vez, aí dentro do registro do paciente, selecionar uma das datas em que o exame foi realizado (lembrando que o paciente pode ter

sido acompanhado por vários anos), e dentro desta data, selecionar uma imagem de cada vez (e cada exame podia ter várias imagens diferentes) para então o aplicativo exportar apenas aquela imagem - o que era inviável em face de termos a intenção de obtermos milhares de imagens para pesquisa (esse volume é um pré-requisito para os algoritmos de *machine learning* terem um bom desempenho na análise de padrões).

Para vencermos essa etapa procuramos na documentação do sistema onde as imagens ficariam armazenadas (normalmente imagens são armazenadas em uma pasta específica do sistema operacional e não dentro do banco de dados). Achamos o parâmetro que identificava a pasta das imagens. Como não tínhamos acesso à essa pasta da rede solicitamos à TI do HOSAG este acesso porém aí nos deparamos com mais um obstáculo inesperado. A própria TI do hospital não tinha acesso àquela pasta. Ela ficava em um servidor na nuvem que era administrado por uma empresa terceira (algo bem comum hoje em dia nas organizações). Solução: a TI abriu um chamado para a empresa terceira pedindo acesso ao servidor. Este é um tópico que merece atenção dos gerentes de projeto pois pode não ser lembrado no momento de planejamento do cronograma e pode comprometer significativamente o prazo de um projeto.

3.4. Extração dos dados

Vencidos os obstáculos para acesso aos dados, o trabalho se focou na criação de *scripts* para seleção de informações - precisávamos de imagens em quantidade equiparada de diagnóstico positivo e negativo para alimentar os modelos de inteligência artificial. Também precisamos nos preocupar com a anonimização dos dados para garantir que nenhuma informação sensível seria compartilhada, mas também sem perder o relacionamento entre imagens \times diagnóstico \times idade dos pacientes.

3.5. Armazenamento

Para poder haver o compartilhamento de dados foi escolhido uma solução *cloud* de mercado (Google). O desafio aqui é garantir o adequado cumprimento de medidas de acesso restritivas para que somente alunos designados pela UDESC pudessem ter acesso de leitura e apenas pelo período que for adequado ao projeto de pesquisa relacionado.

3.6. LGPD

A Lei Geral de Proteção aos Dados reforça a responsabilidade das entidades envolvidas no projeto quanto à proteção e privacidade dos dados mas há exceções que define que o tratamento de dados pessoais, com destaque para informações sensíveis, poderá ser realizado para estudos por órgão de pesquisa, mantendo, sempre que possível, o anonimato dos dados pessoais. No contrato de cooperação não houve explicitamente uma cláusula a respeito da privacidade dos dados com os envolvidos no projeto, mas os mesmos possuíam ciência quanto às responsabilidades por se tratar de dados sensíveis.

Referências

- Conlon, R., Saheb, H., and Ahmed, I. I. K. (2017). Glaucoma treatment trends: a review. In *Canadian Journal of Ophthalmology*, volume 52, pages 114–124. Elsevier B.V.
- Dorion, T. (2002). *Manual de exame do fundo de olho*. Manole, 1st edition.
- Mantravadi, A. V. and Vadhar, N. (2015). Glaucoma. In *Primary Care - Clinics in Office Practice*, volume 42, pages 437–449. W.B. Saunders.