

Título: Predição de Alto Custo para direcionar buscas ativas em Atenção Primária Digital dentro do contexto de Saúde Corporativa

Introdução:

O Grupo 3778 é uma *HealthTech* que fornece soluções inovadoras de gestão de cuidados em saúde para empresas e operadoras de planos de saúde, com serviço de Atenção Primária Direta, baseado em ações de coordenação e navegação de enfermagem.

Para prestar semelhante cuidado, é importante estabelecer critérios objetivos para discriminar, dentro da população, as pessoas que possuem maior risco. É a partir dessa atribuição de risco que a oferta de cuidados é planejada, de modo que seja proporcional à medida da necessidade dos pacientes, sendo assim um processo equânime. *(BRASIL. Ministério da Saúde. Classificação de Risco e Vulnerabilidades na Atenção Primária à Saúde. Brasília: Ministério da Saúde, 2013)*

No entanto, identificar quais pacientes devem ser priorizados para ações de coordenação é um desafio, pois existem inúmeras condições possíveis a se considerar, além disso os dados de sinistro representam de forma incompleta o quadro de saúde do paciente. De todo modo, a solução adotada comumente no mercado de saúde é selecionar pacientes com base em seus altos custos históricos, em geral considerados como os top-10% pacientes mais caros de uma carteira de pacientes. Essa abordagem, entretanto, mostra-se falha, tanto porque os pacientes de alto custo não necessariamente permanecem custando mais, quanto porque a oportunidade de prevenir seus custos foi perdida, apenas olhando dados históricos.

Objetivo:

O objetivo deste trabalho foi desenvolver uma nova abordagem que permitisse prever os altos custos em um horizonte de doze meses e ordenar essas previsões para identificar as pessoas que devem ser colocadas em coordenação preferencialmente. Essa abordagem é baseada na construção de um modelo de predição de alto custo, sendo aplicada em um cenário real como ferramenta auxiliar para realizar buscas ativas de pacientes na atenção primária.

Metodologia:

A metodologia adotada neste estudo baseia-se apenas na utilização de dados de sinistro, que são dados extraídos das faturas de contas médicas apresentadas pelas operadoras após o processamento das cobranças dos procedimentos realizados pelos prestadores.

Para enriquecer esses dados, foi realizada a harmonização de códigos de procedimento e a identificação, com base em heurísticas, dos tipos de atendimentos realizados em um mesmo contexto. Esses tratamentos, por exemplo, permitem identificar, de forma mais consistente, a ocorrência de tempos de internação, tipos de internação (cirúrgica ou clínica) e condições de internação (urgência ou eletiva).

Além disso, para auxiliar o modelo, foram criados outros atributos a nível de paciente, e ao longo do tempo evitando vazamento de informações, onde foram inferidas condições de saúde a partir dos dados de sinistro. Essa inferência conta com cerca de 400.000 regras de inferência e identifica cerca de 40 condições de saúde. Para validar essas regras, foi feita uma

comparação com rótulos atribuídos por médicos em evolução de prontuário, considerados como padrão-ouro. A título de exemplo, a inferência para a doença diabetes teve 16,67% de sensibilidade e 96,40% de especificidade.

A partir desses atributos foi treinado um modelo de *gradient boosting* de regressão implementado pelo *CatBoost*, reservando como conjunto de teste 30% dos pacientes da base de dados. A predição de custo do modelo foi utilizada para ordenar os pacientes do maior para o menor em uma janela de 12 meses.

Para avaliar o modelo, selecionamos os top-10% pacientes com maior custo predito pelo modelo e comparamos com os top-10% pacientes com maior custo real no conjunto de teste, utilizando a precisão como métrica da classificação.

Resultados:

O modelo desenvolvido atingiu cerca de 46% de precisão na identificação dos pacientes com alto custo. Comparado a modelos da literatura, publicados em artigos (*Shaochong Lin, Qingpeng Zhang, Frank Chen, Li Luo, Lei Chen, Wei Zhang, Smooth Bayesian network model for the prediction of future high-cost patients with COPD, International Journal of Medical Informatics, Volume 126, 2019; Osawa, I., Goto, T., Yamamoto, Y. et al. Machine-learning-based prediction models for high-need high-cost patients using nationwide clinical and claims data. npj Digit. Med. 3, Volume 148, 2020.*), o modelo desenvolvido neste estudo apresentou uma performance comparável na predição de alto custo. Em termos de interpretação, ao utilizar o modelo, foi possível identificar cerca de 4,6 a cada 10 pessoas que se tornarão de alto custo nos próximos doze meses, em comparação com apenas cerca de 3 a cada 10 pessoas identificadas pela estratégia clássica de ordenação, que se limita a ordenar os pacientes segundo seu histórico de custo nos últimos 12 meses.

Conclusão:

A implementação desse modelo de predição de alto custo em um sistema de atenção primária digital permitiu automatizar o processo de busca ativa, por um sistema de mensageria, que contata de forma automática pacientes elegíveis através de mensagens de Whatsapp. Essa implementação aumentou a capacidade de buscas ativas da equipe de enfermagem.

O modelo está em constante evolução e deve passar a agregar dados de outras naturezas, para além das contas médicas, como dados de prontuário, atestados, exames ocupacionais, dentre outros.

Essa abordagem demonstra um avanço na gestão da saúde corporativa, proporcionando uma melhor alocação de recursos e equidade do cuidado. Além disso, estudos iniciais de custo efetividade, utilizando o modelo de diferença em diferenças em múltiplos tempos (*Callaway, B. and Sant'Anna, P.H.C. (2021) 'Difference-in-Differences with multiple time periods', Journal of Econometrics, 225(2), pp. 200–230. Available at: <https://doi.org/10.1016/j.jeconom.2020.12.001>.*), apontam uma tendência de redução de custo per capita por mês para os pacientes contemplados por essa estratégia, comparados àqueles que não foram alvo da implementação.

Gregório Rodrigues

Grupo 3778

Manager

Email : gregorio.rodrigues@3778.care

Arthur Reys

Grupo 3778

Data Scientist

Email : arthur.reys@3778.care

Henrique Pickler da Silva

Grupo 3778

Data Scientist

Email : henrique.silva@3778.care

Dacyara Oliveira

Grupo 3778

Epidemiologist

Email : dacyara.oliveira@3778.care