

Monitoramento e impacto assistencial de participantes preditos com alto potencial de variação nas despesas assistenciais: uma aplicação de inteligência artificial na saúde suplementar

Frank Ney Sousa Lima¹, Carlos Wilson Gomes de Barros¹, Antônio Cipriano Neto¹, Bruna Matter dos Santos¹, Wanderley Akira Shiguti¹, Weverton Vieira da Silva Rosa¹.

¹ *Caixa de Assistência dos Funcionários do Banco do Brasil (CASSI) – Brasília/DF*

1º Lugar Prêmio UNIDAS 2025 – Categoria Filiadas

Recebido: 15/01/2026 **Aprovado:** 01/02/2026 **Publicado:** 06/02/2026

Conflitos de interesse: Os autores declaram não haver conflitos de interesse relacionados a este estudo.

Financiamento: Este trabalho não recebeu financiamento específico de agências públicas, comerciais ou sem fins lucrativos.

Autor correspondente: Frank Ney Sousa Lima, Cassi Sede

INTRODUÇÃO

O crescimento das despesas assistenciais na saúde suplementar tem motivado a busca por soluções inovadoras que combinem inteligência analítica e gestão populacional. Diversos fatores explicam essa tendência, como o envelhecimento da população — resultado da queda das taxas de fecundidade — e a elevação da expectativa de vida ao nascer, estimada em 75,5 anos em 2022 (IBGE, 2023) ¹.

É amplamente reconhecido que os padrões de consumo de recursos de saúde variam significativamente com a idade, especialmente quando associados à crescente prevalência das doenças crônicas não transmissíveis (DCNT), em especial as de natureza degenerativa, fatores que somados contribuem para um aumento na demanda por assistência à saúde (SANTOS, TURRA, NORONHA, 2018) ².

Nesse contexto, compreender o perfil populacional que mais consome serviços e identificar, com antecedência, os participantes com maior potencial de impactar as despesas assistenciais permite uma atuação preventiva e a gestão proativa desses grupos, favorecendo a sustentabilidade do sistema.

Em 2021, a operadora de saúde analisada neste estudo registrou um aumento aproximado de 12% em suas despesas assistenciais. Uma análise detalhada revelou que menos de 1% dos participantes sinistrados — aqueles que utilizaram algum serviço de saúde no período — foram responsáveis por aproximadamente 70% da variação total dessas despesas. Esse pequeno grupo, posteriormente denominado Classe 1, apresentava características em comum, como idade avançada, múltiplas comorbidades e alta frequência de utilização de serviços. Após o período analisado, constatou-se ainda que cerca de 40% desses participantes havia falecido, revelando a gravidade das condições de saúde enfrentadas.

Diante desse cenário, surgiu uma pergunta central: seria possível prever, com alguma antecedência, quais participantes apresentariam padrão semelhante de utilização de recursos aos da Classe 1? Esse questionamento motivou o desenvolvimento de um modelo preditivo baseado em machine learning, com o objetivo de antecipar cenários e orientar ações de cuidado mais personalizadas.

O presente estudo descreve a construção e aplicação desse modelo, assim como os

desdobramentos práticos observados após mais de um ano de monitoramento dos participantes denominados “preditos” para Classe 1. A iniciativa representa uma abordagem estratégica de gestão em saúde, que integra inteligência analítica ao planejamento assistencial, favorecendo intervenções mais eficazes, maior vínculo com a atenção primária e uma atuação voltada à sustentabilidade do sistema e à qualidade de vida dos participantes.

OBJETIVOS

Predizer participantes com alta probabilidade de impactar, significativamente, a variação futura das despesas assistenciais e fornecer informações para o monitoramento contínuo da população predita.

MÉTODOS

O desenvolvimento da modelagem preditiva teve como ponto de partida a análise retrospectiva do padrão de consumo assistencial dos participantes que mais impactaram o aumento das despesas da operadora. Foram considerados 48 meses de histórico (de agosto de 2019 a julho de 2023), com cálculo da variação individual das despesas em janelas móveis de 12 meses.

Os participantes que mantiveram o plano ativo durante todo o período e apresentaram as maiores variações nas despesas foram classificados como classe 1. Como os períodos eram móveis, um mesmo participante poderia ser incluído mais de uma vez, mas os registros duplicados foram removidos, resultando em 2.269 participantes únicos nessa classe.

Para prever novos participantes com potencial de impacto semelhante, foi utilizada uma abordagem de modelagem supervisionada por aprendizado de máquinas, técnica que integra o campo da inteligência artificial por permitir que algoritmos aprendam padrões a partir de dados históricos e realizem estimativas sobre novos casos. O objetivo era identificar, com alta probabilidade, indivíduos com maior chance de se tornarem classe 1 no futuro.

A construção do modelo exigiu também um grupo de comparação: a classe 0, formada por participantes com uso assistencial estável e baixo impacto sobre as despesas, igualmente expostos ao período analisado. Como essa classe era numericamente muito superior à classe 1, observou-se um forte desbalanceamento — situação comum em modelos preditivos e que pode comprometer o desempenho do algoritmo. Para contornar essa limitação, aplicou-se a técnica de undersampling, com seleção aleatória de uma amostra considerada estatisticamente representativa da classe 0 (2.500 participantes), conforme proposto por Oliveira e Grácio (2005) ³. Essa abordagem permitiu equilibrar as proporções entre os grupos e melhorar a capacidade preditiva do modelo. Ao final, o conjunto de dados utilizado para a modelagem preditiva foi composto por 4.769 registros, sendo 70% destinados ao treinamento do modelo e 30% ao teste.

Ao todo, foram analisadas 61 variáveis explicativas, envolvendo aspectos demográficos, presença de doenças crônicas, diagnósticos registrados (CID-10), uso de serviços ambulatoriais e hospitalares, entre outros. Após análise de multicolinearidade e avaliação da importância das variáveis, 20 foram mantidas por apresentarem maior poder preditivo.

A modelagem foi implementada em Python, no ambiente Jupyter Notebook, e incluiu a comparação entre diferentes algoritmos de machine learning: Regressão Logística, Random Forest, LightGBM, LogitBoost e XGBoost — modelo baseado em árvores de decisão que obteve o melhor desempenho. Seus indicadores apontaram sensibilidade de 80%, precisão de 73% e acurácia de 76% (grau de acerto geral do modelo).

Após o ajuste do modelo, ele foi aplicado a uma base recente de escoragem. Adotou-se uma abordagem conservadora, classificando como preditos de classe 1 apenas os participantes cuja probabilidade estimada foi igual ou superior a 90%, o que resultou na identificação de 8.081 participantes com alto potencial de impactar as variações nas despesas futuramente.

Como o modelo foi treinado com dados históricos de participantes de alto custo (classe 1) e de baixo custo (classe 0), ele passou a reconhecer, por meio das variáveis analisadas, características comuns a cada grupo — como idade, presença de comorbidades e intensidade no uso de serviços — e estimar, de forma automatizada,

o risco de novos participantes desenvolverem padrões semelhantes.

Essa capacidade de antecipação permitiu ações mais precisas e oportunas na gestão em saúde, pois após a predição, foi iniciado um processo sistemático de monitoramento mensal desses 8.081 participantes preditos, principalmente quanto a utilização da rede própria de Atenção Primária à Saúde e a evolução das despesas conforme o uso da rede credenciada de prestadores. Essa integração entre inteligência artificial e gestão populacional tem contribuído para estratégias mais direcionadas, com foco na prevenção de agravamentos, fortalecimento da atenção primária e maior eficiência no uso dos recursos assistenciais.

RESULTADOS

A aplicação do modelo preditivo resultou na identificação de 8.081 participantes com alto potencial de impactar as futuras variações nas despesas assistenciais. Esses indivíduos, denominados preditos como classe 1 representaram 1,44% da população total analisada. O perfil desses participantes chama atenção pela presença expressiva de pessoas idosas (96% com 60 anos ou mais), alta prevalência de condições crônicas (76% diagnosticados e/ou preditos por modelagem estatística para diabetes, hipertensão e/ou dislipidemia) e incidência significativa de neoplasias (29%). Pouco mais da metade dos participantes preditos era composto por mulheres (53%).

Após a realização da predição, esse grupo passou a ser monitorado mensalmente, possibilitando uma visão ampliada do uso da rede credenciada e dos serviços próprios da operadora. Entre os principais achados, observou-se que, apesar do alto risco estimado, a despesa per capita dos participantes preditos apresentou estabilidade nos meses seguintes à realização do estudo. Nos doze meses anteriores à predição, a despesa per capita dos participantes classe 1 mostrou um aumento de 45% e, um ano após a disponibilização do estudo, houve uma queda de 19%. Paralelamente, poucos meses após a realização da predição, 12% dos participantes eram acompanhados na Atenção Primária e haviam recebido atendimento por um médico de família. Nos doze meses subsequentes, esse percentual aumentou para 40%. Embora não se possa afirmar a existência de uma relação direta entre a predição e a tendência de estabilidade das despesas, os dados sugerem que o monitoramento sistemático e o

olhar assistencial mais direcionado podem contribuir para um uso mais equilibrado dos recursos.

Outro aspecto relevante é a forte presença de doenças crônicas não transmissíveis e condições complexas de saúde entre os preditos, o que reforça a necessidade de estratégias integradas de cuidado, especialmente junto a população idosa. O envelhecimento, que por si só já representa um fator de risco para o agravamento clínico e maior utilização de recursos, torna ainda mais urgente a adoção de modelos de gestão centrados nas necessidades dessa faixa etária.

A alta proporção de participantes preditos como classe 1 com idade entre 80 e 89 anos (41%) e a importância da variável idade na modelagem sinalizam um ponto de atenção para o futuro. Estudos como o de Santos, Turra e Noronha (2018) ² já alertavam para o impacto crescente das DCNT entre idosos e sua relação direta com o aumento das despesas. Além disso, quase metade dos preditos crônicos já apresentavam diagnóstico de doença cardiovascular — principal causa de morte no Brasil — o que reforça a importância de intervenções precoces nesse grupo.

Em consonância com a literatura e as experiências anteriores da própria operadora, a atuação da Atenção Primária em Saúde, por meio do monitoramento de crônicos e da gestão compartilhada do cuidado, pode ser uma aliada fundamental para retardar o agravamento dos quadros clínicos e reduzir internações evitáveis. O cuidado centrado nos idosos, especialmente, exige uma abordagem além do controle de indicadores clínicos: é necessário acolher a complexidade desses indivíduos, garantir continuidade no acompanhamento e promover uma visão longitudinal da saúde.

O monitoramento dos preditos possibilitou também maior articulação entre as áreas da empresa, desde as Equipes de Atenção Primária até núcleos assistenciais especializados, contribuindo para o fortalecimento de ações intersetoriais e gestão populacional ativa. Esse tipo de integração é fundamental diante do perfil prevalente desses participantes, que muitas vezes convivem com múltiplas condições crônicas, fragilidade funcional e risco elevado de desfechos adversos.

Cabe destacar que, embora a predição por si só não represente uma intervenção direta, pode servir como importante gatilho para ações direcionadas, mais oportunas e centradas no indivíduo. No contexto do envelhecimento populacional acelerado e da

busca por maior sustentabilidade do sistema, essa capacidade de antecipar cenários críticos representa um avanço estratégico.

CONCLUSÕES

Os achados deste estudo apontaram caminhos promissores para uma gestão mais estratégica e cuidadosa da saúde suplementar. Ao combinar técnicas de inteligência artificial com um acompanhamento mais próximo dos participantes, foi possível não apenas identificar grupos prioritários, mas, também, orientar ações mais direcionadas e alinhadas às necessidades reais dos indivíduos.

As unidades regionais conheceram mais detalhadamente o perfil da população predita, empreenderam ações específicas conforme a necessidade do público e já se observa uma recomposição na curva de crescimento das despesas desta parcela estudada.

O estudo desenvolvido reforça o valor da análise preditiva como ponto de partida para práticas mais resolutivas, que integram dados e cuidado. Mais do que antecipar riscos, trata-se de criar oportunidades para intervir no tempo certo — com acolhimento, continuidade e foco na pessoa.

REFERÊNCIAS

1. IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Tábuas Completas de Mortalidade para o Brasil – 2022, p. 1-57, 2023. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/3097/tcmb_2022.pdf. Acesso em: 20 jun. 2025.
2. SANTOS, S. L.; TURRA, C. M.; NORONHA, K. Envelhecimento populacional e gastos com saúde: uma análise das transferências intergeracionais e intrageracionais na saúde suplementar brasileira. *Revista Brasileira de Estudos de População*, v. 35, n. 2, p. e0062, 2018. Disponível em: <https://www.scielo.br/j/rbepop/a/Mh8tvhW3PtJQWK7WmKSF5bp/#>. Acesso em: 26 jun. 2025.
3. OLIVEIRA, E. F. T.; GRÁCIO, M. C. C. Análise a respeito do tamanho de amostras aleatórias simples: uma aplicação na área de Ciência da Informação. *DataGramaZero*, 6(3). 2005. Disponível em: https://def.unir.br/uploads/36315994/arquivos/2_1847729620.pdf. Acesso em: 22 jun. 2025.