

MODELO PREDITIVO DO DIABETES – A CLASSIFICAÇÃO DE BENEFICIÁRIOS POR MEIO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA SUPERVISIONADO

Frank Ney Sousa Lima; Gurgel A L A; Oliveira D F; Neto A C; Rios E S; Barros C W G; Barros J F B; Nascimento D C

CASSI – Caixa de Assistência dos Funcionários do Banco do Brasil

OBJETIVOS: Com o desafio de produzir e sistematizar informações sobre as condições de vida e saúde dos beneficiários, gerando conhecimento para orientar a escolha e a aplicação de estratégias de cuidado, a operadora de autogestão em saúde desenvolveu a metodologia de predição para o Diabetes, doença com alta morbimortalidade prematura por conta das suas possíveis complicações. O foco da elaboração do modelo preditivo era estimar, por meio de algoritmo de aprendizagem de máquina, a probabilidade de um determinado beneficiário ser diabético com base no padrão de consumo assistencial, e a partir disso conhecer a prevalência estimada dessa doença na população da operadora.

MÉTODOS: Foi utilizado o método de *Machine Learning* (ML) supervisionado, usando a base de dados do período de agosto de 2016 a março de 2020 das informações dos beneficiários que fazem uso dos serviços próprios da operadora, que dispõe da informação assistencial sistematizada no Prontuário Eletrônico do Paciente (PEP), e das Guias de Utilização de procedimentos do consumo assistencial dos beneficiários nos serviços credenciados. Assim, foi possível buscar padrões de comportamentos realizados por diabéticos, sistematizando-os em um conjunto de variáveis, como idade e sexo, além de exames específicos e outras informações assistenciais. Com isso, as etapas realizadas foram de treino com balanceamento de dados com adição de novas observações (*oversampling*), teste com aplicação de quatro algoritmos, para identificar a melhor performance dos modelos, sendo definido como “campeão” o de Regressão Logística. O tratamento dos dados foi realizado por meio das soluções tecnológicas *SAS Enterprise Guide* e *Python*, e a modelagem, a comparação de modelos e o *scoring* a partir do *software SAS Viya*. A taxa de especificidade do método foi de 90,2%, de acurácia de 89,7% e de sensibilidade de 56,5%. A modelagem seguiu uma proposta conservadora, com a probabilidade de ser diabético igual ou acima de 85%.

RESULTADOS: Inicialmente, a partir dos diabéticos identificados nos registros do PEP, foi calculada uma prevalência de 5,2%, mas após a utilização do algoritmo de classificação, os diabéticos preditos foram computados, resultando em uma estimativa de prevalência de 8,7% da população da operadora com 18 anos ou mais. Esse percentual é próximo do apresentado em pesquisas de saúde nacionais: 7,7% (Pesquisa Nacional de Saúde 2019) e 7,4% (Vigitel 2019).

CONCLUSÕES: O estudo mostrou-se útil para prever a probabilidade do diabetes e orientar a elaboração de medidas para monitoramento, prevenção ou diminuição do seu impacto sobre os beneficiários. Por fim, foi possível constatar que o uso do ML e sua assertividade poderia ser aprimorado e extrapolado para outras condições de saúde, o que já vem sendo desenvolvido na operadora com a produção de modelos preditivos para outras condições crônicas.

Brasília, DF 30 de setembro de 2022

Antonio C. Neto

Barros J. F. B.

Frank Ney S. Lima

Erica Rios