

MODELAGEM PREDITIVA DE DEPRESSÃO EM PACIENTES DIAGNÓSTICADOS COM COVID-19: UTILIZAÇÃO DA TÉCNICA DE REGRESSÃO LOGÍSTICA

Pedro Vitor Bernardes Brandão¹

INTRODUÇÃO

No início do ano de 2020, o novo coronavírus (COVID-19) assolou o mundo e foi definido pela Organização Mundial de Saúde (OMS) como uma emergência de saúde pública (Marin et al., 2020). Diante do surgimento de uma pandemia, aumenta-se a atenção para os cuidados com a saúde, principalmente para a saúde mental.

Segundo Ramos e amigos (2020), as doenças psíquicas têm uma larga prevalência em pandemias, sendo que as principais sequelas psiquiátricas no período pós pandêmico é a ansiedade e a depressão.

De acordo com pesquisa realizada pela Faculdade de Medicina da Universidade Federal de Minas Gerais com um grupo de 35 pessoas que foram contaminadas por COVID-19, um mês após a alta, os usuários relataram dor e impactos no estado emocional com sintomas de ansiedade e

depressão, além de fraqueza nos braços, fadiga e falta de ar.

Frente ao cenário da investigação da depressão como uma das sequelas causadas pela infecção pelo novo coronavírus, encontrar meios que possam atuar de forma preventiva no tratamento dos beneficiários se tornou um dos grandes objetivos do mercado.

A literatura vem apontando o quanto a adoção de modelos para identificação de doenças e riscos de forma antecipada e tratamento precoce reduz os custos operacionais. Mas, o desafio dos modelos de predição com o objetivo de prevenção vai além de categorizar os fatores de risco, dando a eles seus devidos pesos e medidas. Consiste também em ajustar o foco, muitas vezes distorcido, do risco percebido versus o risco real (CARVALHO, 2013).

Tendo em vista este contexto, o objetivo deste estudo é utilizar a técnicas de regressão

¹ pedro.bernardesebrandao@gmail.com – +55 (31) 99120-0889
Rua Severo Belém, 91, Belém, Contagem, Minas Gerais, Brasil – 32015-120

logística para prever a ocorrência da depressão em pacientes que foram diagnosticados com a COVID-19.

MODELO UTILIZADO PARA CLASSIFICAÇÃO: REGRESSÃO LOGÍSTICA

Quando se deseja prever a classificação de certos atributos baseado em outros atributos, a regressão logística é comumente levantada como meio para realização desta classificação, utilizando-se de variáveis dicotômicas como variável resposta, representadas por exemplo, 1 ou 0 (GONZALES, 2018).

Neste estudo, a variável resposta utilizada foi adoecimento por depressão (CID – F32), para a qual 0 indica o não adoecimento e 1 indica o adoecimento pelo desfecho.

No modelo logístico, a relação entre a probabilidade da ocorrência do desfecho é dada pela seguinte função sigmoide:

$$P(Y = 1|X) = \frac{e^{\mu}}{1 + e^{\mu}} \quad (1)$$

na qual μ é uma função que relaciona as variáveis preditoras e tem formulação:

$$\mu = \beta_0 + \sum_{i=1}^{p-1} \beta_i X_i + \epsilon \quad (2)$$

METODOLOGIA

Base de dados

Os dados utilizados neste trabalho foram provenientes do Sistema de Informação Hospitalar do Sistema Único de Saúde (SIH-SUS).

Para realização das análises exploratórias dos dados e consequentemente propor o modelo preditivo, foi realizado um recorte temporal na base de julho de 2020 até julho de 2021.

A base abriga as informações do estado de São Paulo (Brasil) e ao total, possui 2,5 milhões de registros de procedimentos hospitalares realizados no SUS.

Pré-processamento das bases de dados

Para realizar a modelagem preditiva, foram selecionadas da Base as variáveis: código do usuário, data do atendimento, procedimento principal, Classificação Internacional de Doenças - CID, idade e sexo.

A variável de interesse neste trabalho foi a ocorrência do nosso desfecho, que estava na variável CID. As demais variáveis foram suporte no tratamento das bases para que fosse possível identificar os beneficiários que tiveram o desfecho e selecionar a sua utilização durante todo o período.

Identificação e histórico do usuário

Após processadas as bases, foi realizada a coleta das informações dos beneficiários que tinham o registro de depressão no final do período e feita uma análise retrospectiva em um ano de utilização, a fim de verificar qual é o padrão de procedimentos realizados e diagnósticos recebidos desses beneficiários antes de adoecerem por depressão.

Nesta análise, a ideia foi criar uma escada que mostrasse qual é o caminho percorrido por cada usuário até o adoecimento.

Montagem da base final

Após identificado os procedimentos e diagnósticos dos usuários antes de adoecerem por

depressão, foram criadas variáveis que apresentasse o histórico assistencial dos usuários no SUS durante o período de um ano. Desta forma, os registros dos beneficiários foram compilados em uma única linha, contendo as informações assistenciais de julho de 2020 a junho de 2021.

Finalizando o processo de estruturação da base, foi criada uma variável indicadora que atribui 1 para os usuários que tiveram depressão em julho de 2021, mas não tiveram anteriormente e 0 para os usuários que não tiveram depressão em nenhum momento do recorte temporal.

Ambiente Computacional

Para a realização da modelagem preditiva no R, foi utilizado a função glm para o modelo logístico com função de ligação "Logit".

A base final foi dividida em treino e teste, sendo 75% das observações para treino e 25% para teste.

Após criada a base de treino, foi realizado o pré-processamento nos seguintes passos:

1- Criação da regra indicadora;

2- Em seguida, as observações “na” das variáveis categóricas foram consideradas como desconhecidas;

3- As observações “na” das variáveis numéricas foram substituídas pela mediana das informações;

4- Foram excluídas do modelo as variáveis correlacionadas;

5- As variáveis categóricas foram transformadas em variáveis indicadoras (dummies).

Após o pré-processamento, os dados foram treinados e, em seguida, foram preditos os resultados utilizando-se a base de teste.

RESULTADOS E DISCUSSÃO

População estudada

Por meio da análise dos dados, foi possível observar que as mulheres representam 50,35% dos registros no período, enquanto os homens são responsáveis por 49,65% dos registros.

Como aponta a literatura as mulheres têm uma utilização dos serviços de assistência à saúde superior que os homens, porém os dados mostraram uma utilização muito próxima entre homens e mulheres.

Em termos gerais, a base apresenta um perfil etário mais envelhecido, em que a utilização da população jovem representa 21,6% contra 40,65% da população com 60 anos ou mais.

Perfil dos procedimentos e diagnósticos dos usuários com depressão

Em relação aos procedimentos realizados pelos usuários com Depressão, destaca-se o tratamento de infecção pelo coronavírus, diagnóstico ou atendimento de urgência em clínica médica, tratamento de pneumonias ou influenza (gripe) e tratamento de outras doenças bacterianas. Além disso, os pacientes que desenvolveram transtornos depressivos realizaram procedimentos como parto cesariano e parto normal.

Dentre os CIDs apresentados é possível observar que cerca de 16% dos registros se referem a infecção por coronavírus (B34.2), 11% relacionados a parto (O80.0, O80.9, O82.0, O82.1 e O82.9) e 1% relacionado a Broncopneumonia não especificada (J18.0).

Medidas de qualidade dos modelos

Os dados de treino abrigavam 75% da base final, o que representa 8.603 informações, sendo que 0,2% dos usuários possuíam depressão. A base de teste corresponde aos 25% restante da base total (2.868), sendo que 0,2% dos usuários possuíam depressão.

Em relação as medidas de qualidade, o modelo teve 0,9958 de acurácia, ou seja, os resultados da predição estão bem próximos do valor real.

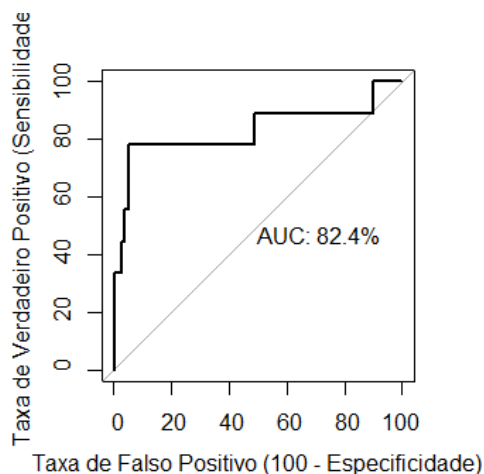
O modelo teve uma alta especificidade (0,99), ou seja, tem uma boa previsão de usuários que o teste deu negativo e que o usuário de fato não apresentava a doença.

O modelo não teve sensibilidade (0), ou seja, teve dificuldade em identificar usuários

que o teste deu positivo e que o usuário de fato apresentou a doença.

Baseando-se nos resultados encontrados foi gerada a curva ROC que é apresentada no gráfico 1 para avaliar a qualidade da técnica.

Gráfico 1 – Curva ROC do modelo de regressão logística



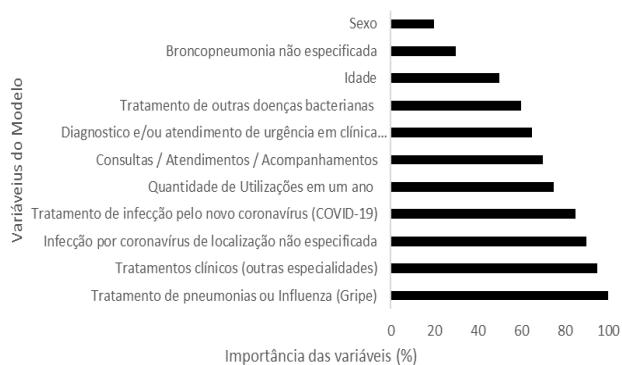
Fonte: Elaboração Própria

O modelo de regressão logística teve uma performance em termos da área abaixo a curva ROC de 82,4%. Na literatura não existe um consenso sobre qual é o valor de AUC que determinaria uma boa capacidade de discriminação de um critério de classificação. Porém, Hosmer e colegas (2013) fornecem um critério de avaliação que segundo ele o modelo aqui proposto

possui excelente poder de discriminação.

Em termos da importância das variáveis para prever a depressão, conforme apresenta o gráfico 2, a quantidade de vezes que realizou tratamento de pneumonias ou influenza (gripe) em um ano foi a variável mais importante. Em seguida a quantidade de vezes que realizou tratamentos clínicos em um ano e depois a quantidade de vezes que foi diagnosticado com COVID-19 em um ano.

Gráfico 2 – Importância das variáveis



Fonte: Elaboração Própria

CONCLUSÃO

Diante da pandemia do COVID-19, o mercado de saúde suplementar vem buscando meios para contornar o aumento dos gastos associado as sequelas

provocadas pelo novo coronavírus. Os modelos preditivos se destacam principalmente pela identificação dos beneficiários que possuem maior propensão a adoecerem e possibilita que as operadoras articulem políticas internas para impedir ou retardar o avanço das doenças.

A baixa sensibilidade está associada principalmente ao desbalanceamento dos dados, em que apenas 0,2% da população foi diagnosticada com depressão. Espera-se que trabalhando com um volume maior de dados e uma prevalência maior da doença na base de estudo melhore a sensibilidade do modelo.

Para os próximos trabalhos espera-se testar novos modelos para prever a depressão em pacientes diagnosticados com COVID-19, além de verificar se a qualidade das previsões aumenta utilizando-se de bancos em que a prevalência da doença seja maior.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBUQUERQUE, Ceres. **A situação atual do mercado da saúde suplementar no Brasil e apontamentos para o futuro.**

2007. Disponível:
<<https://www.scielo.br/j/csc/a/jXwhKzH5MtFjLS4h7WdMy8m/?lang=pt>>.
Acesso em: 06/01/2022

HEIDMANN, Ivonete et al.,
**PROMOÇÃO À SAÚDE:
TRAJETÓRIA HISTÓRICA DE
SUAS CONCEPÇÕES.** 2006.
Disponível:
<https://www.scielo.br/scielo.php?pid=S010407072006000200021&script=sci_abstract>. Acesso em:
11/01/2022

GONZALES, Leandro. **Regressão
logística e suas Aplicações.** 2018.
Disponível:
<<https://monografias.ufma.br/jspui/bitstream/123456789/3572/1/LEANDRO-GONZALEZ.pdf>>. Acesso em:
16/01/2022

MARTINS, Andrigo. **Aplicação de
Análise de Risco de Crédito com
o uso das Técnicas de Regressão
Logística e Árvores de Decisão.**
2020. Disponível:
<<https://repositorio.ufmg.br/handle/1843/35524>>. Acesso em:
23/01/2022

MENDES, Antonio et al.,
Assistência pública de saúde no

**contexto da transição
demográfica brasileira:
exigências atuais e futuras.** 2012.
Disponível:
<<https://www.scielo.br/j/csp/a/YqW3NNYWrvmFWfvkSfmLgpj/abstract?lang=pt>>. Acesso em: 06/01/2022

OHTOSHI, Claudia. **Uma
comparação de regressão
logística, árvore de classificação
e redes neurais: Analisando
dados de crédito.** 2003.
Disponível:
<<https://www.teses.usp.br/teses/disponiveis/45/45133/tde-20210729132841/en.php>>. Acesso em:
17/01/2022