

## ASPECTOS SOCIOAMBIENTAIS COMO PREDITORES DE UM MODELO DE ÁRVORE DE CLASSIFICAÇÃO (CART®) PARA OS REGISTROS DE ÓBITOS E RECUPERADOS POR COVID-19 NO ESTADO DE MINAS GERAIS

Esther Ferreira de Souza<sup>1</sup>  
Fabrício Pelizer Almeida<sup>2</sup>

Saúde Ambiental

### *Resumo*

A atenção global com o cenário de pandemias infecciosas graves considera a vulnerabilidade social como fator determinante na disseminação, especialmente em função da precariedade de serviços como saneamento básico e acesso a água potável. Nestes termos, o objetivo deste trabalho é compor um modelo de árvore de classificação (CART®), para registros de óbitos e recuperados por COVID-19, considerando preditores socioambientais no período de 04 de março à 24 de junho de 2020, para estado de Minas Gerais. Os dados de pacientes recuperados e óbitos confirmados para COVID-19 foram obtidos junto ao sítio da Secretaria de Estado de Saúde de Minas Gerais (SES/MG). Os municípios foram agrupados por regiões de planejamento da SEPLAN/MG, enquanto dados referentes ao saneamento básico municipal foram obtidos no Painel Saneamento Brasil, junto ao Instituto Trata Brasil. Utilizou-se o software Minitab v. 19, o pacote Análise Preditiva e a função CART® classificação, com a seleção de uma fração de 30% das linhas para o teste na composição do modelo estatístico. Os resultados indicam nos primeiros nós, a relevância da presença de alguma comorbidade, associada à faixa etária maior que 52 anos como determinante no obituário. As regiões do Rio Doce, Central e Zona da Mata, sem comorbidade ou não informado, classificam registros de óbitos, aprofundando no sentido de estratos de saneamento básico descritos como inadequados. Conclui-se que o modelo de árvore de classificação (CART®), pode apoiar ações estratégicas do governo estadual, sem dissociar os aspectos sociais e ambientais como fatores determinantes de respostas para a COVID-19.

Palavras-chave: Algoritmo CART; Saneamento Básico; Pesquisa Qualitativa; Epidemiologia.

<sup>1</sup> Graduanda em Engenharia Ambiental, Universidade de Uberaba (Uniube), [estherriquel.mus@gmail.com](mailto:estherriquel.mus@gmail.com).

<sup>2</sup> Prof. Dr. Universidade de Uberaba (Uniube) – Instituto de Tecnologia, [fabricio.almeida@uniube.br](mailto:fabricio.almeida@uniube.br).

## INTRODUÇÃO

Estudos que combinam dados de pandemias respiratórias infecciosas graves já relatadas globalmente fornecem evidências convincentes, de que o aumento do risco em uma população pode ser amplamente impulsionado por fatores de vulnerabilidade social, e agravados pelas desigualdades econômicas e acesso desproporcional aos cuidados de saúde, especialmente serviços de saneamento básico (NASSIF-PIRES et al., 2020). Além é claro, da desproteção da saúde da população idosa e descaso quanto aos fatores associados à comorbidade, tão recorrentes em conglomerados urbanos dos países pobres e em desenvolvimento (GUAN et al., 2020; MELO et al., 2020).

Especificamente para o atual cenário global da COVID-19, um dos desafios é o de elucidar um modelo classificatório de variáveis correspondentes à evolução dos casos notificados da doença. Portanto, objetiva-se com este trabalho desenvolver um modelo de árvore de classificação (CART<sup>®</sup>), para registros de óbitos e recuperados por COVID-19 no período de 04 de março à 22 de junho de 2020, para estado de Minas Gerais.

## METODOLOGIA

O estudo apoia-se no levantamento de notificações oficiais de casos de pacientes recuperados\* e óbitos confirmados para COVID-19 junto ao sítio da Secretaria de Estado de Saúde de Minas Gerais (SES/MG), no período de 04 de março à 24 de junho de 2020. Os dados secundários considerados nesta pesquisa comportam ainda variáveis do paciente como sexo, idade, presença ou não de comorbidade e município de residência do paciente.

Optou-se por compor o modelo de classificação das variáveis, a partir do agrupamento dos municípios obedecendo-se a divisão em Regiões de Planejamento estado de Minas Gerais adotada pela Secretaria Estadual de Planejamento e Gestão do Estado de Minas Gerais (SEPLAN), e não individualizados. Os dados referentes ao saneamento básico municipal foram obtidos no Painel Saneamento Brasil, junto ao sítio do Instituto

---

\* Casos confirmados de COVID-19 que receberam alta hospitalar e/ou cumpriram isolamento domiciliar de 14 dias sem intercorrências.

Trata Brasil – ITB (2020). No Quadro 1 estão dispostas as variáveis adotadas no modelo de classificação, o tipo e descrição, atentando-se àquelas que sofreram alguma conversão.

Quadro 1. Variáveis, tipos e descritivos dos preditores e resposta do modelo de classificação da árvore de decisão – CART<sup>®</sup>

Variável	Tipo	Descrição
'Evolução dos Casos de Covid-19'	Resposta Categórica	<b>Óbito; Recuperado.</b>
'Idade'	Preditor Contínuo	Registro informado pelo paciente.
'Sexo'	Preditor Contínuo	<b>Feminino (1); Masculino (2).</b>
'Comorbidade'	Preditor Categórico	<b>Sim; Não; Não Informado.</b>
'Região de Planejamento'	Preditor Categórico	Agrupamento dos municípios correspondente às 10 (dez) regiões do estado de MG. <sup>1</sup>
'Classe do Serviço de Saneamento'	Preditor Categórico	<b>Inadequado</b> (prevalece o Índice de Atendimento sem Coleta e sem Tratamento); <b>Pouco Adequado</b> (prevalece o Índice de Atendimento com Coleta e sem Tratamento); <b>Adequado</b> (prevalece o Índice de Atendimento com Coleta e com Tratamento).

<sup>1</sup>Fundação João Pinheiro (2010), autarquia vinculada à Secretaria Estadual de Planejamento e Gestão do Estado de Minas Gerais (SEPLAN).

Foram empregadas as regras de divisão dos nós e sub-nós, do algoritmo CART<sup>®</sup>, utilizando-se o critério de Gini, conforme Breiman, et al. (1984). O método de validação foi obtido a partir de um conjunto teste automático, com a seleção de uma fração de 30% das linhas, com uma base de geração de dados aleatório. O teste e treinamento do modelo demonstraram saídas idênticas tanto pelo critério do custo mínimo quanto de K erros-padrão. As simulações para a construção da árvore de decisão foram realizadas no pacote **Análise Preditiva**, função **CART<sup>®</sup> classificação**, do software Minitab v.19.

## RESULTADOS E DISCUSSÃO

Foram satisfeitas as condições de qualidade e elevado poder e especificidade do teste *versus* treinamento para o evento resposta “Óbito”, em torno de 93,8% e 96,9%, respectivamente, e proximidade das predições para o evento “Recuperado”, 85,9% (treinamento) e 85,9% (teste), conforme disposto na Tabela 1.

Tabela 1. Matriz de confusão para predição (treinamento e teste)

Classe real	Classe predita (Treinamento)				Classe predita (Teste)			
	Contagem	Óbito	Recuperado	%Correto	Contagem	Óbito	Recuperado	%Correto
<b>Óbito</b> (Evento)	543	526	17	<b>96,9</b>	242	227	15	<b>93,8</b>
<b>Recuperado</b>	13426	1899	11527	<b>85,9</b>	5715	807	4908	<b>85,9</b>
<b>Todos</b>	13969	2425	11544	<b>86,3</b>	5957	1034	4923	<b>86,2</b>

Os resultados no diagrama de árvore demonstram a efetiva condição de presença de alguma comorbidade de pacientes (não definido o fator de risco) como o primeiro desdobramento de classificação de obituários. No segundo estrato, a faixa etária ‘maior que 52 anos’ determina a classificação de obituário e, não havendo comorbidade, o público com idade ‘maior que 60 anos’ responde melhor para recuperação da doença (Figura 1).

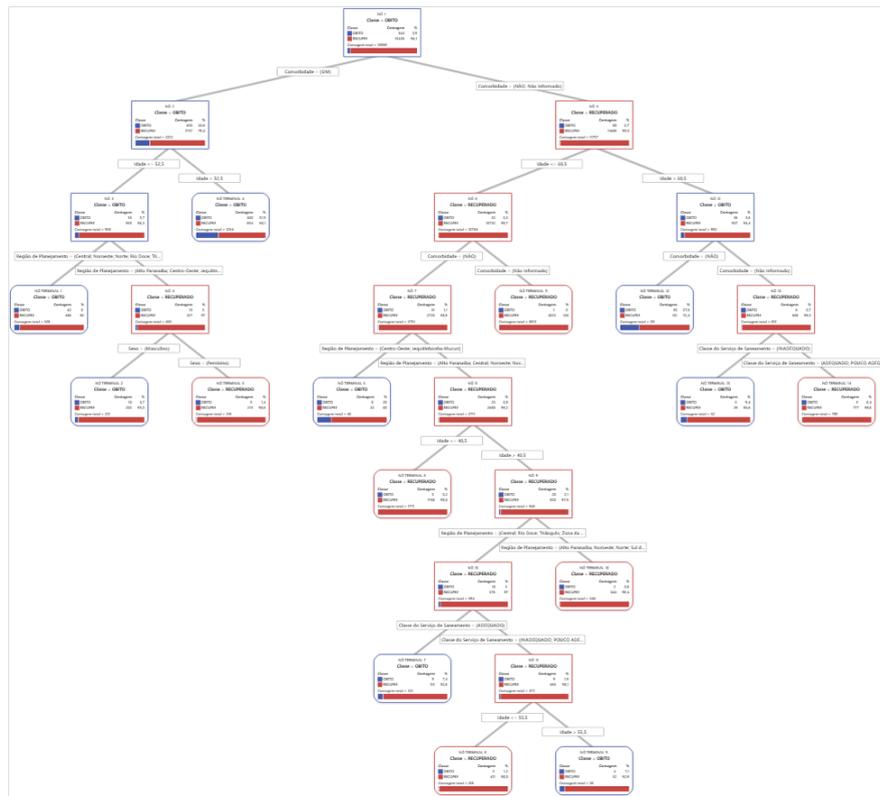


Figura 1. Diagrama de árvore ótima (menor custo ótimo relativo), com 6 nós terminais.

As regiões Noroeste, Norte, Rio Doce e Central classificam maior número de óbitos no terceiro nível de nós para o público ‘maior que 52 anos’ e com indicação de comorbidade no obituário. Não havendo comorbidade, ou sem informação, há maior expressão das regiões Centro-Oeste e Jequitinhonha-Mucuri para a resposta categórica. O grupo

masculino ramifica maior óbito em nós finais, entre os residentes nas regiões do Alto Paranaíba e Centro-Oeste, faixa etária ‘menor que 52 anos’ e registro de comorbidade.

Os nós terminais, para os óbitos por COVID-19 encerram a faixa etária ‘maior que 53 anos’ e comorbidade, e entre os idosos (acima de 60 anos) em condições inadequadas de prestação do serviço de saneamento básico. Em faixas etárias em torno de 40 anos, os nós finais associam as regiões do Rio Doce, Central e Zona da Mata e a inadequação do serviço de saneamento como determinantes para o maior registro de obituários pela doença.

Os resultados reforçam a importância de estratégias com enfoque em grupos sociais vulneráveis, sem dissociá-los dos aspectos econômicos e ambientais locais, com a devida atenção ao serviço de saneamento básico em regiões mais críticas do estado.

## CONSIDERAÇÕES FINAIS

Os resultados obtidos neste trabalho podem orientar através do modelo de árvore de classificação (CART®) ações estratégicas do governo estadual em parceria com os municípios, observando-se as classes de maior risco para a COVID-19, associadas aos inúmeros desafios sociais, econômicos e ambientais das diferentes regiões do estado.

## REFERÊNCIAS

- BREIMAN, L. et al. Classification and Regression Trees. Boca Raton, Florida: Chapman & Hall/CRC. 1ª ed. 1984. 368p.
- GUAN, W.; et al. Comorbidity and its impact on 1,590 patients with COVID-19 in China: A Nationwide Analysis. European Respiratory Journal. v. 55, n. 6, p. 2-56, 2020.
- FUNDAÇÃO JOÃO PINHEIRO (FJP). Produto Interno Bruto (PIB) dos municípios de Minas Gerais, 2010. Disponível em: <http://www.fjp.mg.gov.br/>. Acesso em jul. de 2020.
- INSTITUTO TRATA BRASIL. Painel Saneamento Brasil, Ranking do Saneamento, 2019. Disponível em: <http://www.tratabrasil.org.br/>. Acesso em jul. de 2020.
- MELO, R. R. P. B de.; et al. Desafios no acesso à água e saneamento básico no Brasil e o controle da COVID-19. Revista Augustus, Rio de Janeiro, v.25, n. 51, p. 281-293, 2020.
- NASSIF-PIRES, L.; et al. Pandemic of Inequality. Public Policy Brief. Levy Economics Institute of Bard College. n. 149, 2020. 16p.