

## ÁRVORE DE CLASSIFICAÇÃO PARA OS REGISTROS DE CASOS NOTIFICADOS DE DENGUE ASSOCIADOS À VARIÁVEIS CLIMÁTICAS NO MUNICÍPIO DE UBERLÂNDIA/MG

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

Saúde Ambiental

### *Resumo*

A proposição de modelos que incluem um razoável número de variáveis contínuas e categóricas capazes de delinear o comportamento de séries epidemiológicas tem sido objeto de estudo e investigação de várias linhas de trabalho, que consideram o conhecimento computacional e a urgência de tratativas no âmbito da saúde pública. O objetivo deste trabalho é o de desenvolver um modelo de árvore de classificação (CART®), incluindo registros climatológicos e os casos notificados de dengue, no intervalo de 2015 e 2019 para o município de Uberlândia/MG. Os registros climáticos foram obtidos junto ao SADMET/INMET, enquanto os casos notificados de dengue solicitados a SES/MG, no período de janeiro de 2015 a dezembro de 2019. Para a composição do modelo de árvore de classificação foi necessário categorizar duas variáveis (estacionais e faixa de amplitude média de temperatura semanal) e mantendo-se os meses do ano e a pluviosidade total estacional como variáveis contínuas. Os procedimentos estatísticos foram realizados no software Minitab v. 19, com o uso do pacote Análise Preditiva e a função CART® classificação, respeitando-se as regras para a divisão de nós e sub-nós, e o critério do custo mínimo. A árvore de decisão indicou forte intervenção estacional (outono, inverno e primavera) e períodos específicos do ano (segundo semestre) para como critério de decisão para a elevada notificação (evento de resposta). Conclui-se que a atenção dos programas de vigilância sanitária no município deve extrapolar os períodos críticos biológicos do vetor, considerando a extensão anual de infecção e de registros de casos da doença.

Palavras-chave: Algoritmo CART; Pesquisa Qualitativa; Sazonalidade; Sustentabilidade.

<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

A correlação entre fatores climáticos adversos e a ocorrência de eventos epidemiológicos graves e expansivos tem sido o objetivo de vários estudos globais simulatórios em torno da dinâmica de disseminação geográfica e sazonal da doença (CARNEIRO et al., 2017; CORTES, et al., 2018; BISHT et al., 2019). Há um consenso crítico quanto à associação de arboviroses, especificamente aqui a notificação por dengue no Brasil, e a deficiência no suprimento de água tratada e esgotamento sanitário, manejo inadequado de resíduos sólidos, alterações climáticas bruscas e o status socioeconômico local deficiente (RODRIGUES et al., 2012; FERREIRA et al., 2017).

Um grande desafio é compor um modelo inteligente de alerta e preditivo, e que favoreça o planejamento e gestão pública municipal da enfermidade, e nesta direção, o trabalho de Lowe et al., (2016) aponta um modelo precoce baseado em previsões categóricas de dengue abrangendo várias microrregiões brasileiras.

Neste sentido, objetiva-se com este trabalho desenvolver um modelo de árvore de classificação (CART<sup>®</sup>), composto por variáveis climáticas e registros semanais de casos notificados de dengue, no intervalo de 2015 e 2019 para o município de Uberlândia/MG.

## METODOLOGIA

O estudo apoia-se no levantamento de dados secundários de séries históricas climatológicas semanais – precipitação total (mm), temperatura média (°C), junto à SADMET/INMET, referente à estação climatológica automática instalada no município de Uberlândia/MG (A507), e os casos notificados de dengue, junto ao sítio da Secretaria de Estado de Saúde de Minas Gerais (SES/MG) no mesmo município, obedecendo-se o calendário semanal epidemiológico – no período de 01 de janeiro de 2015 à 31 de dezembro de 2019.

Foi necessário alinhar os dados semanais epidemiológicos e os registros climáticos para o mesmo período de referência. Inclusões quantitativas como a incidência da doença no município (relativa à população censitária, e equivalente à 100.000 pessoas), e

amplitudes térmicas semanais (diferença entre as temperaturas máximas e mínimas) foram admitidas no banco de dados. Entretanto, esta última variável foi descrita no modelo como faixa binária categórica, tal como o efeito sazonal, dado pelas estações do ano. A variável dependente “Situação da Incidência Anual”, categoriza os casos notificados da doença, adotando-se o evento de resposta “Alta Incidência”, conforme descrito no Quadro 1.

Quadro 1. Tipos e descritivo dos preditores e variável resposta do modelo de classificação da árvore de decisão – CART<sup>®</sup>

Variável	Tipo	Descrição
Situação da Incidência Anual 'Situ. Anual'	Resposta Categórica	<b>Baixa</b> ( $0 < y' \leq 150$ ); <b>Alta</b> ( $y' > 150$ )
Pluviosidade Total Estacional 'Pluv. T. Est.'	Preditor Contínuo	Obtido à partir dos dados de pluviosidade da estação corrente.
'Mês.'	Preditor Contínuo	$m = 1, 2, 3, \dots 12.$
Faixa de Amplitude Semanal 'Faixa. Ampl.'	Preditor Categórico	<b>Amena</b> (Se a amplitude média semanal $\leq$ amplitude média estacional); <b>Elevada</b> (Se a amplitude média semanal $>$ amplitude média estacional).
'Estação'	Preditor Categórico	Verão: ' <b>ve</b> '; Outono: ' <b>ou</b> '; Inverno ' <b>in</b> '. Primavera: ' <b>pr</b> ';

Foram empregadas as regras de divisão dos nós e sub-nós, do algoritmo CART<sup>®</sup>, utilizando-se o critério de Gini para a divisão dos dados, conforme Breiman, et al. (1984). Para fins de teste foram excluídos aleatoriamente 25% dos dados, porém de modo balanceado quanto aos anos e semestres de registros. 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

A atenção com a qualidade do teste *versus* treinamento no trabalho é atendida pela elevada proximidade dos percentuais de acerto nos dois procedimentos para as duas classes reais e preditas categorizadas (Tabela 1). Também as sensibilidades satisfatórias do teste (91,2%) e treinamento (91,7%) permitem inferências acerca do modelo.

Tabela 1. Matriz de confusão e poder do teste e treinamento

Classe real	Classe predita (Treinamento)				Classe predita (Teste)		
	Contagem	Alta	Baixa	%Correto	Alta	Baixa	%Correto
Alta (Evento)	181	166	15	91,7	165	16	91,2
Baixa	49	0	49	100,0	5	44	89,8
Todos	230	166	64	93,5	170	60	90,9
<b>Estatísticas</b>					<b>Treinamento (%)</b>		<b>Teste (%)</b>
Taxa de positivo verdadeiro (sensibilidade ou poder)						91,7	91,2
Taxa de positivo falso (erro tipo I)						0,0	10,2
Taxa de negativo falso (erro tipo II)						8,3	8,8
Taxa de negativo verdadeiro (especificidade)						100,0	89,8

Os resultados demonstram a expressiva condição temporal-sazonal na determinação dos registros de casos de dengue no município de Uberlândia/MG. O caminho decisório percorrido pelo evento de resposta “Alta Incidência” confere maior preocupação no segundo semestre (primeiro nó ótimo) e, em sequência, a identificação de “Baixa Incidência” em estratos com registros pluviométricos mais elevados (Figura 1).

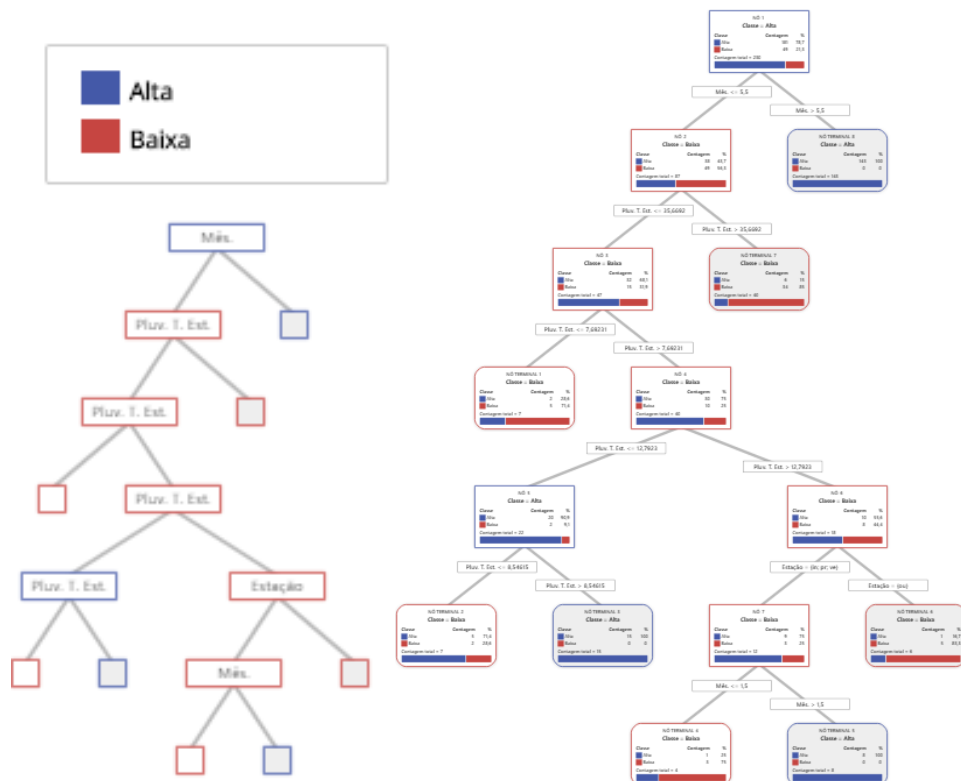


Figura 1. Diagrama de árvore ótima (com divisão de nós e detalhada), com 8 nós terminais.

Os nós seguintes reforçam ainda mais o efeito sazonal e pluviométrico como determinantes para o padrão crítico da variável resposta, principalmente após 45 dias do início da estacional (inverno, primavera), e faixas pluviométricas medianas (próximo à 8,54 mm). Nestes termos, a gestão pública municipal de considerar o efeito residual de infecção do período mais quente e chuvoso (primeiro trimestre), coincidindo com o eventual afrouxamento dos trabalhos de vigilância sanitária, em detrimento a outros desafios da saúde pública típicos do período seco. É preciso ainda, ter atenção quanto à adaptabilidade do vetor às condições urbanas adversas, especialmente quanto aos criadouros.

## CONSIDERAÇÕES FINAIS

Os resultados deste trabalho permitem traçar ótimas perspectivas de aplicação das árvores de decisão, com a seleção e ajustes do algoritmo CART<sup>®</sup> na gestão de casos notificados de dengue, em uma proposta de implementação de estratégias preditivas.

## REFERÊNCIAS

- BISHT, B. et al. Influence of environmental factors on dengue fever in Delhi. *International Journal of Mosquito Research*, v. 6, n. 2, p. 11-18, 2019.
- BREIMAN, L. et al. *Classification and Regression Trees*. Boca Raton, Florida: Chapman & Hall/CRC. 1<sup>a</sup> ed. 1984. 368p.
- CARNEIRO, M. A. F. et al. Environmental factors can influence dengue reported cases. *Rev. Assoc. Med. Bras.*, São Paulo, v. 63, n. 11, p. 957-961, 2017.
- CORTES, F. et al. Time series analysis of dengue surveillance data in two Brazilian cities. *Acta Tropica*, v. 182, p. 190-197, 2018.
- FERREIRA, D. A. da C. et al. Meteorological variables and mosquito monitoring are good predictors for infestation trends of *Aedes aegypti*, the vector of dengue, chikungunya and Zika. *Parasites & Vectors*, v. 10, n. 78, p.1-11, 2017.
- LOWE, R. et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *Cambridge: Elife Sciences Publications Ltd*, v. 5, p. 1-18, 2016.
- RODRIGUES, N. C. P. et al. Temporal and Spatial Evolution of Dengue Incidence in Brazil, 2001-2012, *PLoS One*, v. 11, n. 11, p. 1-12, 2016.