ANÁLISE PREDITIVA DE CASOS DE DENGUE EM SENADOR CANEDO-GO UTILIZANDO APRENDIZADO DE MÁQUINA

Luís Eduardo Morais dos Santos (Discente Apresentador)
Diogo Gonçalves Dias (Orientador)

RESUMO

A dengue representa, historicamente, um dos mais graves desafios de saúde pública no Brasil e no Estado de Goiás, exigindo o desenvolvimento de ferramentas analíticas capazes de antecipar surtos epidêmicos e orientar decisões estratégicas governamentais. Este trabalho de Iniciação Científica e Tecnológica teve como objetivo principal modelar e prever o número de casos de dengue no município de Senador Canedo (GO), utilizando dados históricos de notificações de saúde e aplicando técnicas avançadas de estatística e de aprendizado de máquina. A metodologia adotou um fluxo computacional rigoroso, estruturado no ambiente Google Colaboratory com a linguagem Python e a biblioteca Pandas. A fase de engenharia de dados envolveu a ingestão da base original, conversão temporal dos registros de notificação, filtragens para remoção de inconsistências e a agregação da evolução diária, semanal e mensal dos casos, incluindo o cálculo de médias móveis de sete dias para suavização de flutuações de registro. Com a base sanada, o núcleo do estudo consistiu na implementação, calibração e comparação empírica de cinco abordagens de modelagem preditiva aplicadas sobre um conjunto de teste contendo os últimos 24 meses da série histórica. Os modelos avaliados foram: SARIMA(1,0,0)(0,0,0)[12], SARIMA(1,1,1)(1,1,1)[12], Suavização Exponencial Tripla (Holt-Winters Aditivo), o algoritmo Prophet (Meta) e o Naïve Sazonal, adotado como benchmark mínimo de precisão. O desempenho preditivo foi rigorosamente quantificado por meio de cinco métricas de erro padronizadas: Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE), Erro Quadrático Médio (MSE), Erro Percentual Absoluto Médio (MAPE) e Coeficiente de Determinação (R^2). A análise exploratória confirmou um padrão sazonal de altíssimo risco concentrado no primeiro semestre do ano, especificamente entre os meses de fevereiro e abril, coincidindo com o período mais chuvoso da região central de Goiás, enquanto a janela de menor transmissão ocorre de julho a outubro. Os resultados preditivos evidenciaram que os modelos com componente sazonal explícito apresentaram notável superioridade na antecipação do início dos surtos. Destacou-se o modelo SARIMA(1,1,1)(1,1,1)[12], que obteve o menor RMSE e o maior R^2 no conjunto de teste, demonstrando excelente aderência aos dados reais. O modelo Prophet também exibiu grande versatilidade na decomposição de tendências plurianuais, embora exija calibração cuidadosa para não subestimar picos extremos. Conclui-se, portanto, que a integração de métodos econométricos clássicos e abordagens de ciência de dados fornece uma base científica indispensável para a vigilância epidemiológica municipal. A adoção desse pipeline computacional permite prever o volume de casos com meses de antecedência, viabilizando o planejamento de campanhas de conscientização e o dimensionamento de recursos de forma mais proativa, rápida e embasada em evidências científicas sólidas.

Palavras-chave: Dengue; Aprendizado de Máquina; Séries Temporais; SARIMA; Saúde Pública.