

Núcleo de Avaliação: Núcleo II

Área temática: Ciência da Computação

Área do Conhecimento: Ciências Exatas e da Terra

Modelos de predição de casos de Dengue em Mossoró – Uma análise comparativa entre algoritmos de Machine Learn

Mikael Johnatan Azevedo, João Pedro Pereira Frutuoso, Amanda Gondim de Oliveira,
Angélica Félix de Castro, Paulo Gabriel Gadelha Queiroz

A dengue é uma doença infecciosa transmitida pelo mosquito *Aedes aegypti* e representa um dos principais desafios de saúde pública em regiões tropicais, especialmente no Brasil, onde apresenta comportamento endêmico e surtos recorrentes. Esses surtos geram sobrecarga aos serviços de saúde e podem evoluir para quadros graves, destacando a importância de estratégias capazes de prever a incidência de casos com antecedência, permitindo ações preventivas e de controle. Nesse contexto, modelos de Machine Learning têm se mostrado promissores na modelagem de séries temporais epidemiológicas. Este trabalho teve como objetivo avaliar diferentes algoritmos de Machine Learning e Deep Learning na previsão de casos de dengue no município de Mossoró/RN, utilizando dados epidemiológicos e climáticos. Os dados de casos confirmados foram obtidos do Sistema de Informação de Agravos de Notificação (SINAN), enquanto as variáveis climáticas (temperatura média, precipitação e umidade relativa do ar) foram coletadas da estação meteorológica da UFRSA, abrangendo o período de 2010 a 2024. No pré-processamento, valores climáticos faltantes foram preenchidos com a média histórica do mesmo dia ao longo dos cinco anos anteriores, preservando a sazonalidade. Valores extremos foram ajustados pelo método IQR com Winsorização. Em seguida, os dados diários foram agregados para escala mensal e criaram-se lags de 30 e 60 dias dos casos de dengue para capturar tendências temporais associadas ao ciclo de transmissão da doença. Foram avaliados cinco modelos: Random Forest (RF), Support Vector Regression (SVR), Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM) e Convolutional Neural Network (CNN) e o desempenho foi medido por R^2 e MSE. Para verificar diferenças estatísticas entre os modelos, aplicou-se o teste de Friedman, seguido do pós-teste de Nemenyi. Os resultados indicaram diferença significativa no desempenho (p -value = 0,0008), destacando a CNN como superior ao MLP e ao Random Forest. No entanto, não foram observadas diferenças significativas entre CNN, LSTM e SVR, sugerindo desempenhos semelhantes entre esses modelos. Conclui-se que modelos baseados em Deep Learning, especialmente CNN e LSTM, apresentam maior potencial para prever casos de dengue em cenários reais, podendo ser integrados a sistemas de vigilância epidemiológica. Trabalhos futuros incluem a ampliação da previsão para nível de bairros, inclusão de variáveis demográficas e automação da atualização dos dados.

Palavras-chave: Predição, Machine Learn, Dengue, Saúde Pública

Agência financiadora: PIBIC/CNPq

Campus: Mossoró

