

191^a DEFESA DE DISSERTAÇÃO EM ENGENHARIA INDUSTRIAL

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA INDUSTRIAL - PEI

TAIANE CAMPOS BRAGA DE ASSIS



pei@ufba.br



www.pei.ufba.br



@peiufba



@peiufba



PEI TV

Orientadores:

- Prof. Dr. Cristiano Hora de Oliveira Fontes (PEI-UFBA);
- Prof. Dr. Édler Lins de Albuquerque (PEI-UFBA).

Banca Examinadora:

- Prof. Dr. Cristiano Hora de Oliveira Fontes (PEI-UFBA);
- Prof. Dr. Édler Lins de Albuquerque (PEI-UFBA).
- Prof. Dr. Marcelo Augusto Costa Fernandes (UFRN);
- Prof. Dr. Roberto Guardani (USP).

Suplentes:

- Prof. Dr. José Geraldo Andrade Pacheco (UFPE).

Título: IMPUTAÇÃO DE DADOS AUSENTES EM SÉRIES TEMPORAIS DA QUALIDADE DO AR: UM ESTUDO DE CASO PARA A CIDADE DE SALVADOR.

Data: 23 de julho de 2026

Horário: 14h00min

Local: https://conferenciaweb.rnp.br/webconf/pei_epufba

Resumo:

A presença de dados ausentes constitui um dos principais desafios para a utilização de séries temporais de qualidade do ar, podendo comprometer análises ambientais, estudos epidemiológicos e processos de tomada de decisão. Nesse contexto, este trabalho teve como objetivo avaliar diferentes abordagens de aprendizado de máquina para a imputação de dados ausentes em séries temporais de qualidade do ar, considerando padrões reais de ausência observados na rede de monitoramento atmosférico de Salvador BA. Inicialmente, foram analisados os padrões de dados ausentes e as relações entre as variáveis monitoradas por meio de análises de completude, correlação cruzada e definição de variáveis exógenas. Em seguida, foram implementados e comparados os modelos XGBoost, Long Short-Term Memory (LSTM) e Recurrent Neural Network (RNN) para a imputação das concentrações de monóxido de carbono (CO) em uma das estações de monitoramento existentes na cidade de Salvador (estação do Centro Administrativo da Bahia, CAB). A avaliação dos modelos foi realizada por meio das métricas Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação (R^2), complementadas por testes estatísticos de normalidade, comparação e análise de resíduos. Os resultados indicaram que os três modelos apresentaram desempenho satisfatório na reconstrução das concentrações de CO, sendo o XGBoost o modelo com menores erros medianos. Entretanto, de forma geral não foram observadas diferenças estatisticamente significativas entre os resultados de predição. A análise de sensibilidade aos conjuntos de variáveis exógenas demonstrou que cenários contendo outros poluentes atmosféricos apresentaram desempenho superior à utilização exclusiva de variáveis meteorológicas, evidenciando a importância da seleção de variáveis associadas à mesma dinâmica de emissão da variável alvo. Conclui-se que técnicas de aprendizado de máquina constituem uma alternativa robusta para a imputação de dados ausentes em séries temporais da qualidade do ar. Ademais, os resultados evidenciam que a compreensão dos processos físicos e das relações entre os poluentes monitorados constitui fator tão importante para o sucesso da imputação quanto a escolha e aplicação adequada do modelo de predição. A metodologia proposta apresenta potencial de aplicação em diferentes estações de monitoramento e contextos ambientais, sendo a seleção adequada das variáveis exógenas um elemento central para a reconstrução confiável das séries temporais de qualidade do ar analisadas.

Palavras-chave: qualidade do ar; dados ausentes; imputação de dados; aprendizado de máquina; séries temporais; XGBoost; redes neurais recorrentes.