Evaluation of machine learning algorithms and their contribution to the analysis of air quality data and the prognosis of sick building syndrome

Fabián Leonardo Cuesta
Astudillo, Laura Margarita
Rodríguez Peralta
Facultad de Tecnologías de
Información y Ciencias de Datos
Universidad Popular Autónoma
del Estado de Puebla
fabian.cuesta@upaep.edu.mx
lauramargarita.rodriguez01@upaep.mx

Abstract

Indoor air quality is a critical aspect directly impacting the health and well-being of individuals within a building. However, due to air pollution and its effects on occupants, there exists a risk of developing sick building syndrome. The aim of this study is to assess the effectiveness of various machine learning algorithms applied to analyze data related to factors influencing air quality and predict occurrences of sick building syndrome, with the goal of determining the most effective algorithm. A total of 75,213 samples were analyzed, and after training with each machine learning algorithm, the performance metrics of each algorithm were compared. The study concludes that decision trees and Random Forest are the most stable algorithms across different training scenarios for predicting sick building syndrome, achieving metrics ranging between 99% and 100%, respectively.

Keywords

Machine learning, air quality, sick buildings, respiratory diseases, metrics