Data analysis using Machine Learning to determine air quality and environments conducive to COVID19 contagion, lung diseases and 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

Low air quality inside the buildings affects the well-being of its inhabitant. The existence of polluting agents in the environment is a risk to their health and to the existence of the so-called sick building syndrome. The objective of this study is to analyze the data obtained with an intelligent sensor inside a building dedicated to higher education, acquiring information about the components that affect air quality, according to the standards of the Environmental Protection Agency of the United States (EPA), with the purpose of identifying the level of risk of an environment conducive to the contagion of lung diseases, including COVID19, using Machine Learning algorithms. 50,000 samples were taken and trained in the Python programming language through a Support Vector Machine as a risk classifier for sick buildings, reaching an effectiveness result in the model of 95.59%.

Keywords

Machine learning, Air quality, sick buildings, respiratory diseases, covid19