

# Air quality analysis with supervised learning algorithms in Coyhaique, Chile

*Mailiu Díaz Peña<sup>1</sup>, Felipe A. Medina Marín<sup>2</sup>, Ana Karina Maldonado Alcaíno<sup>3</sup>, Dante Cáceres Lillo<sup>4</sup>*

<sup>1</sup>mailiu.diaz@unab.cl, Facultad de Ingeniería, Universidad Andrés Bello

<sup>2</sup>f.medina@uchile.cl, Instituto de Salud Poblacional, Facultad de Medicina, Universidad de Chile

<sup>3</sup>vgtariana@uchile.cl, Instituto de Salud Poblacional, Facultad de Medicina, Universidad de Chile

<sup>4</sup>dcaceres@uchile.cl, Instituto de Salud Poblacional, Facultad de Medicina, Universidad de Chile

Coyhaique is one of the most polluted cities in Latin America, located in the southern region of Chile, with a population density of no more than 7 inhabitants per  $km^2$ . Currently, it is among the three cities with the highest concentrations of particulate matter 2.5 (PM2.5) in the country, mainly due to the use of firewood as a source of energy and heating, negatively impacting the quality of life of those who reside in the area. However, meteorological factors also influence the levels of PM2.5 concentration, so it is relevant to quantify their importance. To do this, we implemented supervised learning algorithms using the data available in the last 10 years from fixed air quality monitoring stations in Coyhaique. We used these models to predict the favorable conditions under which PM2.5 exceeds the Chilean standards, which define a limit of  $20 \mu g/m^3$ . Cross-validation was used with different performance metrics to quantify the accuracy and precision of the fitted models. The evaluation allows for explaining the trend of PM2.5 concentrations and predicting its behavior, which can help decision-makers optimize resource allocation to comply with air quality standards.

**Keywords:** particulate matter, machine learning.