

PRONÓSTICO PRELIMINAR DE MATERIAL PARTICULADO CON REDES NEURONALES Y MEDICIONES DE BAJO COSTO

Ariel F. Scagliotti^{1,2}, Guillermo A. Jorge^{1,2}, Josefina Urquiza^{2,3}, S. Enrique Puliafito^{2,4}, María Florencia Tames^{2,4}
ascagliotti@campus.unqs.edu.ar. Ariel Scagliotti

¹ Instituto de Ciencias. Universidad Nacional de General Sarmiento
² CONICET

³ Facultad Regional Córdoba. Universidad Tecnológica Nacional

⁴ Grupo de Estudios Atmosféricos y Ambientales. Facultad Regional Mendoza. Universidad Tecnológica Nacional

Palabras clave: sensores, aprendizaje automático, calidad del aire

1) INTRODUCCIÓN

El deterioro de la calidad del aire (CA) es uno de los mayores problemas ambientales en todo el mundo (OMS, 2022). Generar nuevo conocimiento al respecto resulta indispensable para aplicaciones regulatorias, de investigación y forenses (Rybarczyk and Zalakeviciute, 2018). Uno de los principales indicadores de la CA son los aerosoles suspendidos en la atmósfera, que cercanos a la superficie suelen ser caracterizados como concentraciones de material particulado (PM). Sin embargo, los equipos requeridos para su monitoreo son extremadamente costosos y las mediciones son generalmente dispersas o inexistentes en algunas regiones (Tian y otros, 2013). Este es el caso de Argentina, que posee por ejemplo escasas estaciones de monitoreo en la Ciudad de Buenos Aires y ninguna en el conurbano bonaerense.

En los últimos tiempos ha aumentado la disponibilidad tanto de sensores de bajo costo para mediciones ambientales como de placas de desarrollo y programación en código abierto, lo que ofrece una alternativa interesante a los equipos de alto costo. Esta tecnología tiene numerosas limitaciones, pero resulta prometedora y la literatura respalda su utilización para la evaluación cualitativa de las concentraciones de partículas en el aire (Lewis y otros, 2018).

Con respecto al modelado de la CA, un enfoque que se ha utilizado con notable éxito es el de Redes Neuronales Artificiales (ANN). Estas pueden ser utilizadas para regresión, clasificación y predicción, e intentan extraer la mayor cantidad de información posible con los datos disponibles (Basheer and Hajmeer, 2000). Resultan una herramienta con mucho potencial para estudiar la CA en regiones con limitada disponibilidad de datos.

Este trabajo presenta el desarrollo de un pronóstico de PM menor a $2,5\mu\text{m}$ de diámetro ($\text{PM}_{2,5}$) con un modelo ANN Perceptrón Multicapa, a partir de mediciones de bajo costo y datos de capa límite atmosférica, en el centro-norte del conurbano bonaerense ($34^{\circ}31'22.8''\text{S}$ $58^{\circ}42'00.3''\text{W}$).

2) METODOLOGÍA

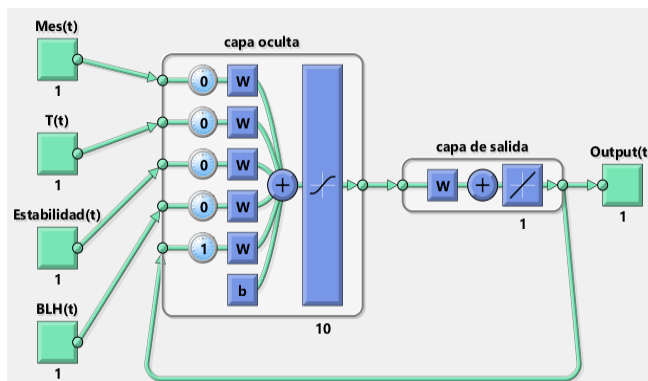
Se utilizaron datos del prototipo de bajo costo para la CA ECAUNGS, que se muestra en la Figura 1 (Scagliotti and Jorge, 2022). Equipo controlado con una placa Arduino y que mide

humedad, presión, temperatura, PM_1 , $PM_{2.5}$ y PM_{10} . Adicionalmente, se descargaron datos de altura de capa de mezcla y condiciones de estabilidad atmosférica de la región en la página de NOAA (NOAA READY, 2021). Se ajustaron las frecuencias de muestreo de todos los datos para trabajar en base horaria, del 06/07/2019 al 01/02/2021.



Se determinó a $PM_{2.5}$ de ECAUNGS como la variable objetivo a pronosticar. Luego, se realizaron ajustes lineales y no lineales entre todas las variables disponibles y $PM_{2.5}$, descartando las que mostraron baja correlación. Con las variables seleccionadas, se entrenó una ANN Perceptrón Multicapa con hiper-parámetros por defecto y 10 neuronas ocultas (Figura 2).

Figura 1. Equipo ECAUNGS instalado



Finalmente, se probaron diferentes reglas de aprendizaje, funciones de transferencia para las neuronas ocultas, métodos de normalización de los datos y cantidades de neuronas ocultas. Se realizaron 10 entrenamientos cada vez y se promediaron los R^2 (coeficientes de determinación) y RMSE (raíz del error cuadrático medio) de cada modelo, para finalmente seleccionar el mejor.

Figura 2. Arquitectura general de la ANN ensamblada. Los predictores seleccionados fueron Mes (mes correspondiente a cada medición), T (temperatura medida por ECAUNGS), Estabilidad (condiciones de estabilidad atmosférica) y BLH (altura de capa de mezcla).

3) RESULTADOS

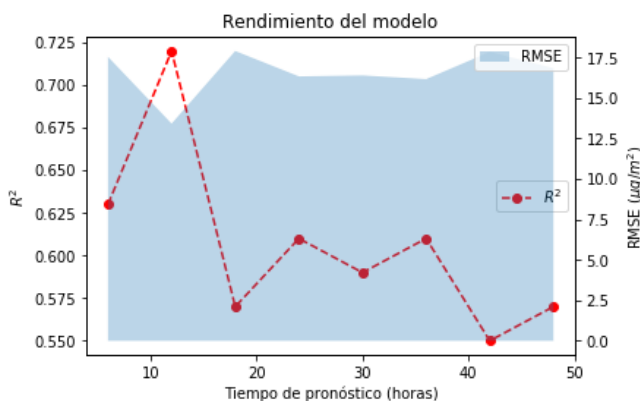
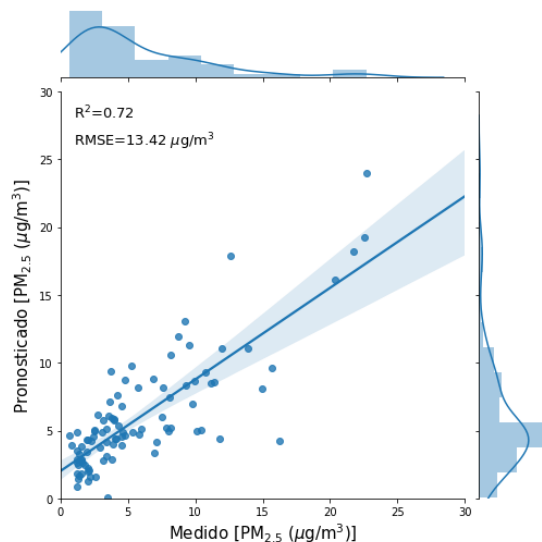


Figura 3. Coeficiente de determinación y RMSE de la salida del modelo respecto a los datos medidos en función de las horas de anticipación para el pronóstico.

Se consiguió el mejor desempeño de pronóstico con una realimentación de la salida con un retraso de 1 h, 10 neuronas ocultas, función de transferencia tangente hiperbólica y regla de aprendizaje de regularización bayesiana. Se observó una mejoría en el pronóstico al realizar un filtro de medianas de 3er orden a la variable objetivo (para el entrenamiento del modelo, luego el test con nuevos datos se realizó con los datos no filtrados) y al normalizar los datos de entrada con un mapeo entre -1 y 1.

Al implementar el modelo bajo diferentes horas de anticipación, se determinó el mejor rendimiento del mismo con un pronóstico a 12 h (Figura 3). Se observó una lógica

disminución en el desempeño a medida que se intenta pronosticar con mayor cantidad de horas de anticipación. Sin embargo, también se observó dentro de esa tendencia general que el modelo logra un mejor ajuste cuando pronostica con múltiplos de 12 h de anticipación (24 h, 36 h, 48 h).



La Figura 4 muestra el ajuste entre los datos pronosticados a 12 h y los medidos por ECAUNGS. Los resultados son prometedores, aunque se observa una gran dispersión que se refleja en un alto RMSE.

Figura 4. Ajuste lineal entre el pronóstico a 12 h y las mediciones reales.

4) CONCLUSIONES

Se obtuvieron buenos resultados a partir del modelado con ANN. La correlación de la salida del modelo con los valores medidos fue alta, con un error en principio aceptable. Sin embargo, se debe considerar que un modelo de tal magnitud necesita ser entrenado y probado con un mayor registro histórico de datos con el fin de considerarlo óptimo en su convergencia y generalización. Este trabajo representa una primera iniciativa en el pronóstico de la CA a partir de mediciones con sensores e instrumentos de bajo costo en la zona de estudio. Se continuarán recabando mediciones y perfeccionando el modelo, agregando más predictores y situaciones particulares para conseguir un entrenamiento más fiable y completo.

5) REFERENCIAS

Basheer, I. A., y Hajmeer, M., 2000: Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1), 3-31.

Lewis, A., Peltier, W. R., y von Schneidmesser, E., 2018: Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications.

Organización Mundial de la Salud: <https://www.who.int/es>. Visitado: 20/03/2022.

Rybarczyk, Y., y Zalakeviciute, R., 2018: Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences*, 8(12), 2570.

Scagliotti, A. F., y Jorge, G., 2022: Métodos de bajo costo para estudio preliminar de material particulado en el centro-norte del conurbano bonaerense. Vol. 33, No. 1, pp. 18-23.

Tian, H., Qiu, P., Cheng, K., Gao, J., Lu, L., Liu, K., y Liu, X., 2013: Current status and future trends of SO₂ and NO_x pollution during the 12th FYP period in Guiyang city of China. *Atmospheric Environment*, 69, 273-280.