

Calibración de un modelo de red neuronal para pronóstico de concentraciones de PM10

Olmos-Guerrero Héctor Antonio ^{1,*} y Rodríguez-González Pablo Tenoch ^{1,2}

¹ Maestría en Ciencias en Ingeniería Ambiental, TecNM, Instituto Tecnológico de Aguascalientes, Av. Adolfo López Mateos Ote. 1801, Bona Gens, Aguascalientes, Aguascalientes, 20256, México.

² Cátedras CONACYT – TecNM campus Aguascalientes, Av. Insurgentes Sur 1582, Ciudad de México, 03940, México.

* Autor de correspondencia: hector.og@aguascalientes.tecnm.mx

Contaminación de agua, suelo y aire (Monitoreo de contaminantes). **Ponencia Virtual.**

Recibido: 16 de junio de 2023 Aceptado: 25 de agosto de 2023 Publicado: 23 de noviembre de 2023

Palabras clave: Redes neuronales artificiales; PM10; Contaminación del aire; Modelado.

Introducción. Uno de los contaminantes que se vigila en la calidad del aire por el daño que puede provocar a la salud son las partículas menores a 10 micrómetros (PM10). El monitoreo y modelado de este tipo de contaminantes es crucial para evaluar riesgos y desarrollar estrategias de mitigación. Se han desarrollado múltiples modelos para generar predicciones de concentración de PM10, algunos tipos son fenomenológicos, estadísticos, aprendizaje automático, sistemas de información geográfica, entre otros. Dentro de los modelos de aprendizaje automático podemos señalar los de redes neuronales artificiales (RNA) que pueden ser usados para generar pronósticos sobre calidad del aire (Agarwal y col., 2020), ya que son capaces de procesar grandes cantidades de datos para encontrar patrones o relaciones entre parámetros. En este trabajo se presenta un prototipo de modelo entrenado con una base de datos de horizonte de tiempo corto de concentración de PM10 de una zona de monitoreo de la ciudad de Aguascalientes.

Materiales y Métodos. Para el entrenamiento del modelo prototipo se utilizaron datos de concentración (PM10) de un mes (2023-05-13 al 2023-06-13) de la estación de monitoreo ubicada en la Secretaría del Medio Ambiente de la ciudad de Aguascalientes. La base de datos utilizada es pública y puede ser consultada en la página del SINAICA (Instituto Nacional de Ecología y Cambio Climático [INECC], 2023). Los datos fueron filtrados antes del entrenamiento. Para el filtrado se desarrolló un *script* en *Python* que elimina lecturas de concentraciones negativas.

El modelo se elaboró usando la librería *Keras* (Chollet y col., 2015) en *Python*. Los parámetros de entrada fueron el día de la semana [0,6], la hora del día [0,23] y el valor de la concentración de PM10 una hora anterior a la actual [0,∞], mientras que la salida es la concentración de PM10 actual [0,∞]. Después del filtrado se obtuvieron 756 registros de cada uno de los parámetros mencionado. Los parámetros de entrada se escalaron utilizando máximos. La base de datos fue separada: 70% entrenamiento, 20% validación y 10% prueba. Los hiperparámetros del modelo fueron; topología 3-100-100-1, tangente hiperbólica como función de activación de las capas internas y lineal a la salida, como función perdida el error medio cuadrático (MSE). El modelo fue entrenado por 100 épocas utilizando el gradiente descendente con momentum (SGD).

La validación se hizo a partir de un análisis de los residuales, homocedasticidad, auto correlación y auto correlación parcial, error medio absoluto (MAE) y cuadrático (MSE), y el coeficiente de determinación (r^2), no confundir con el coeficiente de correlación.

Resultados. La Figura 1 (a) muestra como durante el entrenamiento del modelo la función perdida va disminuyendo con en cada época de entrenamiento, notar que la línea de validación no supera a la del entrenamiento, lo que nos indica que el modelo no está sobre entrenado. La Figura 1 (b), presenta la dinámica de PM10 (datos entrenamiento) y la evaluación del modelo. La escala de esta figura (Figura 1 (b)) esta acotada hasta $200 \mu\text{g}/\text{m}^3$ para poder apreciar el

seguimiento del modelo a los datos reales, ya que en el conjunto de datos de entrenamiento (medición) se tenían algunos valores atípicos altos de concentración, los cuales el modelo no fue capaz de predecir. Los resultados de validación obtenidos del entrenamiento fueron: $r^2 = 0.3753$, $MSE = 974.8962$ y $MAE = 17.8955$. Estos resultados de MAE, MSE y r^2 se encuentran alrededor de valores esperados de un entrenamiento correcto para este tipo de modelos (Fernando y col., 2012).

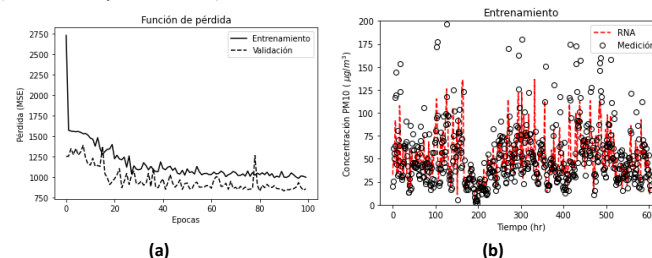


Figura 1. (a) Función perdida durante las épocas del entrenamiento y (b) predicciones horarias del modelo (línea roja) y valores reales (círculos vacíos) de PM10.

Las pruebas de normalidad indicaron que los residuales no siguen un comportamiento normal, pero si se tiene homocedasticidad en los residuales. Las pruebas de autocorrelación mostraron que los residuales no siguen ningún patrón (figuras no mostradas en este trabajo). La evaluación del modelo utilizando los datos prueba generó $r^2 = 0.11104$, $MAE = 22.7253$ y $MSE = 1055.7539$; los cuales eran de esperarse un poco más bajos que los del entrenamiento.

Conclusiones. El prototipo de modelo propuesto muestra resultados satisfactorios a la hora de hacer estimaciones de concentración aplicado en un horizonte de tiempo corto, además que cumple con las condiciones de validación (MAE, MSE, r^2 , homocedasticidad y autocorrelación del error), pero las predicciones del modelo pueden llegar a mejorarse al incorporar algún nuevo parámetro que esté relacionado con los valores atípicos altos que se pueden llegar a encontrar en algunos puntos temporales, como información sobre emisiones y fuentes de área. Esta metodología puede ser utilizada para un conjunto de datos mayor (de varios años) para tener un modelo robusto que pueda dar información más precisa a la hora de hacer pronósticos en cualquier periodo del año.

Bibliografía.

- Agarwal, S., Sharma, S., Suresh, R., Rahman, M. H., Vranckx, S., Maiheu, B., & Batra, S. (2020). Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Science of the Total Environment*, 735, 139454.
- Chollet, F. and other (2015). *Keras*. <http://keras.io>.
- Fernando, H. J., Mammarella, M. C., Grandoni, G., Fedele, P., Di Marco, R., Dimitrova, R., & Hyde, P. (2012). Forecasting PM10 in metropolitan areas: Efficacy of neural networks. *Environmental pollution*, 163, 62-67.
- Instituto Nacional de Ecología y Cambio Climático (INECC). (2023). Sistema Nacional de Información de la Calidad del Aire. Recuperado de <https://sinaica.inecc.gob.mx/>