

Aplicación de modelo XGB para predicción de calidad de aire, Bogotá -Colombia

Gonzalo Ricardo Novoa Fernández^{1,*}

¹Estudiante de Maestría en Ingeniería, Fundación Universitaria Los Libertadores

*Autor de correspondencia: grnovoaf@libertadores.edu.co



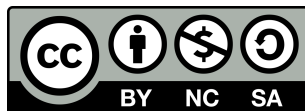
Facultad de Ingeniería y
Ciencias Básicas



Recibido: 4 de julio de 2023

Aceptado: 4 de agosto de 2023

Publicado: 24 de noviembre de 2023



Copyright: ©2023 por los autores. Este artículo es de acceso abierto distribuido bajo los términos y condiciones de Creative Commons Licencia de atribución (CC BY NC SA)
(<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Resumen

El monitoreo y el pronóstico de la calidad de aire es una tarea relevante para la adopción de sistemas de alerta de alerta temprana y medidas de control para reducir o evitar la exposición a aire contaminado y sus efectos en la salud. Metodología: se aplicó el modelo Extreme Gradient Boosting - XGB para la predicción de valores de material particulado PM2.5 y PM10 para la ciudad de Bogotá – Colombia. Se utilizó la información de las concentraciones de los contaminantes provenientes de la Red de Monitoreo de Calidad del Aire de Bogotá para 15 estaciones para el periodo comprendido entre el 1 de enero de 2021 y el 14 de junio de 2023; se realizó el análisis preliminar de los datos y descripción del comportamiento de los contaminantes, posteriormente, se aplicó y validó el modelo XGB. Resultados: se identificaron las concentraciones de valor medio diario para la ciudad de $16.5 \mu\text{g}/\text{m}^3$ para PM2.5 y $32.9 \mu\text{g}/\text{m}^3$ PM10, las mayores concentraciones se registran en el periodo entre 6:00 a.m. y 10:00 a.m.; en el mes de febrero y en el sur occidente de la ciudad. Se realizó el entrenamiento del modelo utilizando el 80% de los datos y la validación con 20% restante identificando que el modelo XGB predice de forma acertada y consistente los valores de las concentraciones de los contaminantes al ser comparado con los datos de prueba, se obtuvo un puntaje de entrenamiento de 0.72 y un error de raíz cuadrada media de 11.73 Conclusiones: el modelo XGB permite la predicción acertada y robusta de las concentraciones de PM2.5 y PM10 con lo cual se puede obtener información complementaria al sistema de pronóstico existente en la ciudad y contribuir al fortalecimiento de las acciones para garantizar la calidad de aire.

Palabras clave: Aprendizaje automático, Calidad de aire, material particulado, Modelo XGB

Como citar este artículo

Gonzalo Ricardo Novoa Fernández. "Aplicación de modelo XGB para predicción de calidad de aire, Bogotá -Colombia", *Revista Apuntes de Ciencia e Ingeniería*, 2, 1, nov, pag 5-14. 2023. doi: [10.37511/apuntesci.v2n1a3](https://doi.org/10.37511/apuntesci.v2n1a3)

1. Introducción

La contaminación del aire es un tema relevante debido a sus impactos en la salud, se estima que el 99 % de la población a nivel mundial respira aire contaminado; así mismo, son importantes los efectos en el ambiente dado que el aire contaminado afecta los ecosistemas y el clima en todo el mundo, siendo una de las causas del cambio climático [1].

La contaminación del aire se debe a la presencia de agentes químicos, físicos o biológicos que modifican las características naturales del mismo. Los principales contaminantes del aire son el material particulado (PM por sus siglas en inglés) y los contaminantes gaseosos como el bióxido de carbono (CO₂), monóxido de carbono (CO), hidrocarburos (HC), bióxido de azufre (SO₂) y óxidos de nitrógeno (NO_x). El material particulado está compuesto por partículas sólidas (ej. sulfato, nitrato, amonio, iones de hidrógeno, carbono elemental, aluminio, silicio, potasio, calcio etc.) y líquidas (ej. Agua). El material particulado se clasifica dependiendo del tamaño de las partículas así: PM₁₀ partículas con diámetros menores a 10 micras, el PM_{2.5} con un diámetro de menor a 2.5 micras y PM₁ partículas ultrafinas con diámetros menores a una micra [2].

El monitoreo de la calidad de aire permite detectar las variaciones de la concentración de los contaminantes del aire y contribuye a la gestión de las autoridades ambientales, el monitoreo requiere inventarios de emisiones (concentraciones de emisiones) y datos meteorológicos [3]. Las medidas de monitoreo de calidad de aire son una herramienta valiosa; sin embargo, existe una demanda creciente de estrategias que permitan realizar pronósticos o predicciones de la calidad de aire ya que dichas estrategias permiten la elaboración de políticas de gestión urbana y de movilidad y de políticas públicas encaminadas a prevenir los posibles efectos en la salud y el ambiente [4]. Las predicciones de calidad de aire en las ciudades es una tarea desafiante debido a que la contaminación del aire depende de la interacción de diferentes factores como las condiciones meteorológicas, la infraestructura urbana y las actividades humanas, las fuentes de emisión y las transformaciones fisicoquímicas de los contaminantes del aire, entre otros [5].

2. Marco teórico

Para el monitoreo de la calidad del aire se utilizan habitualmente dos tipos de métodos de pronóstico: los modelos clásicos de difusión física y los modelos basados en datos [5]. Los modelos clásicos de difusión física pueden proporcionar una interpretación confiable de los procesos físicos involucrados en la dispersión de los contaminantes del aire y en el transporte debido a que son modelos matemáticos que incorporan los datos de dinámica atmosférica, química ambiental atmosférica y datos históricos de contaminación del aire y meteorológicos para calcular la distribución temporal y espacial de los contaminantes [6]. Sin embargo, estos modelos pueden no funcionar bien cuando hay desviaciones en las condiciones meteorológicas o las fuentes de emisión de contaminantes, adicionalmente, no puedan capturar algunas interacciones no lineales complejas entre diferentes factores que afectan la calidad del aire [5]. Son ejemplos de los modelos clásicos de difusión física: el modelo de pluma gaussiana [7], el modelo comunitario de calidad del aire multiescala (CMAQ por sus siglas en inglés) [8], el modelo de química geológica (GEOS-Chem) [9] y el modelo fotoquímico WRF/Chem [10].

Por otra parte, los modelos basados en datos se basan en técnicas estadísticas y de aprendizaje automático, estos modelos pueden capturar relaciones no lineales complejas observadas en los factores que afectan la calidad del aire y se pueden utilizar a gran escala [5], [11]. Los modelos basados en datos se agrupan en modelos estadísticos lineales y modelos estadísticos no lineales. Los modelos estadísticos lineales establecen la relación entre las características relacionadas y la calidad del aire a través de un modelo lineal, son ejemplos de este grupo el Modelo autorregresivo integrado de media móvil (ARIMA) [12], el modelo de regresión de usos del suelo (LUR por sus siglas en inglés) [13], [14], y el modelo de regresión lineal multivariable (MLR) [15]. En el grupo de los modelos estadísticos no lineales, se encuentran, el modelo de máquinas de vectores de soporte (SVM) [16], el modelo de redes neuronales artificiales [17] y el modelo de árbol de decisión [18].



Los modelos de árboles de decisión por su parte incluyen el modelo Extreme Gradient Boosting (XGB), este modelo utiliza árboles de decisión para pronosticar la variable de interés, el algoritmo permite el procesamiento secuencial haciendo uso de crecimiento por niveles en cada iteración, con lo cual aprende automáticamente del error y mejora el desempeño [19]. Este modelo ha sido previamente utilizado como método pronóstico de la calidad del aire [18], [20] y puede presentar algunas ventajas frente a los demás modelos utilizados debido a que proporciona un rendimiento mejorado en términos de corrección del sesgo de pronóstico[21]; asimismo, el modelo trata de manera eficiente los valores faltantes en el conjunto de datos, puede evitar el problema de sobreajuste y aumentar la eficiencia computacional [22].

Bogotá es una ciudad colombiana ubicada a 2600 msnm, que reporta continuamente problemas de calidad de aire por la presencia de material particulado derivada de las actividades industriales y fuentes móviles (vehículos) [23]. La ciudad realiza el monitoreo de la calidad de aire a través de la Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB), la cual está conformada por 18 estaciones de monitoreo fijas y 2 estaciones móviles. Las estaciones registra de forma continua información sobre la concentración de los contaminantes (PM10; PM2.5, SO₂, NO₂, CO, O₃ y black carbon) y las variables meteorológicas (temperatura, radiación solar, humedad relativa, presión barométrica, precipitación, velocidad y dirección del viento) [24].

La ciudad cuenta con un Sistema Integrado de Modelación de Calidad de Aire de Bogotá (SIMCAB) que permite estimar el destino y distribución de concentración de los contaminantes atmosféricos inventariados en la ciudad, el pronóstico permite estimaciones del estado de la calidad de aire para 48 horas a futuro en una resolución espacial de 1 km², la predicción de condiciones meteorológicas emplea el modelo WRF (Weather Research and Forecasting Model), el inventario de emisiones y el modelo fotoquímico que involucra la herramienta CMAQ (Community Multi-scale Air Quality) [25]. Aunque la ciudad cuenta con pronósticos, estos tienen restricciones en cuanto al tiempo que es posible predecir y que no todas las estaciones de monitoreo proveen de información completa y confiable, por lo cual se pueden incorporar y aprovechar las bondades del modelo XGB para proveer de información adicional y complementaria a la suministrada por el SIMCAB que ayude a la toma de decisiones en el monitoreo de la calidad de aire en zonas críticas de la ciudad. Con este propósito el objetivo de este trabajo es aplicar el modelo XGB para la predicción de valores de material particulado PM2.5 y PM10 para la ciudad de Bogotá – Colombia.

3. Metodología

Para el desarrollo de esta investigación se obtuvo la información de las concentraciones de PM2.5 y PM10 por cada hora del día de la Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB) proveniente de 14 estaciones fijas y 1 estación móvil para el periodo comprendido entre el 1 de enero de 2021 y el 14 de junio de 2023 (equivalente 894 días y 23 horas). Para la selección de las estaciones de monitoreo se consideraron los siguientes criterios: a) que la estación de monitoreo estuviera en funcionamiento durante la totalidad del periodo analizado para garantizar que los datos fueran suficientemente representativos y b) que los registros de medición de los parámetros de interés fueran válidos.

Inicialmente se desarrolló el análisis preliminar que incluyó la identificación de valores perdidos y la descripción del comportamiento de las concentraciones diarias del material particulado. Posteriormente, se aplicó y validó el modelo Extreme Gradient Boosting (XGB), para entrenar el modelo se utilizaron el 80 % de los datos disponibles, el 20% restante se utilizó para la validación de los resultados obtenidos. Para verificar la estabilidad del modelo, se seleccionaron datos de diferentes intervalos de tiempo como conjunto de prueba.

Para mejorar el modelo XGB y lograr un rendimiento óptimo se consideraron los parámetros máxima profundidad del árbol (Max_depth=3), número de árboles (n_estimators=1000) tasa de aprendizaje por refuerzo (learning_rate=0.01), los demás parámetros utilizaron sus valores predeterminados. Para el análisis se utilizó el paquete de Python xgboost.



4. Resultados

A continuación, se presentan los resultados obtenidos, inicialmente la descripción del análisis preliminar de los datos y posteriormente los resultados de la aplicación del modelo XGB para la predicción de calidad de aire en Bogotá.

Concentración diaria PM_{2.5}

Al realizar la exploración de los datos de las concentraciones de PM_{2.5} se identificó que las estaciones con menor número de valores faltantes fueron en su orden Ministerio de Ambiente, Bolivia y Fontibón. En contraste la estación de Carvajal es la que mayor porcentaje de valores no válidos registra (Tabla ??).

	Valores cero	Valores faltantes	% Valores faltantes	Total valores no válidos	% Total valores no válidos
Carvajal	3	12007	55.9	12010	55.9
Usaquen	833	4665	21.7	5498	25.6
Mov.Fontibón	11	2936	13.7	2947	13.7
Colina	42	2364	11.0	2406	11.2
Suba	307	1895	8.8	2202	10.3
S.Cristobal	1533	617	2.9	2150	10.0
Kennedy	180	1704	7.9	1884	8.8
P.Aranda	478	1022	4.8	1500	7.0
Tunal	657	690	3.2	1347	6.3
CAR	698	540	2.5	1238	5.8
Guaymaral	391	767	3.6	1158	5.4
Las Ferias	308	688	3.2	996	4.6
Fontibón	64	879	4.1	943	4.4
Bolivia	4	786	3.7	790	3.7
M.Ambiente	362	372	1.7	734	3.4

El promedio por hora de la concentración de PM_{2.5} para la ciudad oscila entre 14.1 y 22.6 $\mu\text{g}/\text{m}^3$, y el valor medio diario de es 16.5 $\mu\text{g}/\text{m}^3$. En el análisis por estaciones de monitoreo se identificó que las mayores concentraciones se registran en la jornada de la mañana entre las 6:00 y 10:00 a.m., posteriormente, se presentan fluctuaciones con valores inferiores a 23 $\mu\text{g}/\text{m}^3$ (ver Fig.1). El mes con mayores concentraciones diarias es febrero. Se destaca que la estación de Carvajal ubicada en el sur occidente de la ciudad reporta los valores más altos. Por el contrario, las menores concentraciones se registraron en la estación el Tunal.

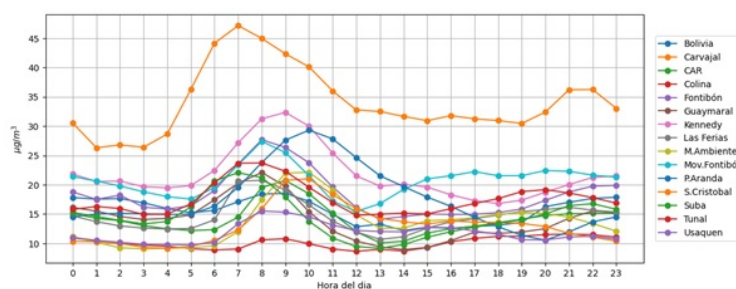


Figura 1: Concentración media de PM_{2.5}. por cada hora del día (enero 2021 - junio 2023)

Concentración diaria PM₁₀

En la exploración de los datos de las concentraciones de PM₁₀ se identificó que las estaciones de Fontibón, Tunal y San Cristóbal son las que presentan mayor porcentaje de valores válidos, así mismo, se destaca que la estación de Carvajal es la que presenta mayor porcentaje de valores no válidos.



	Valores cero	Valores faltantes	% Valores faltantes	Total valores no válidos	% Total valores no válidos
Carvajal	0	11923	55.5	11923	55.5
P.Aranda	0	7075	32.9	7075	32.9
Usaquen	2	4642	21.6	4644	21.6
Colina	6	2770	12.9	2776	12.9
Kennedy	0	2682	12.5	2682	12.5
Mov.Fontibón	1	2617	12.2	2618	12.2
M.Ambiente	0	2501	11.6	2501	11.6
Bolivia	1	1888	8.8	1889	8.8
CAR	1	1791	8.3	1792	8.3
Suba	3	1753	8.2	1756	8.2
Las Ferias	0	1622	7.6	1622	7.6
Guaymaral	1	1249	5.8	1250	5.8
S.Cristobal	34	1109	5.2	1143	5.3
Tunal	21	1084	5.0	1105	5.1
Fontibón	4	954	4.4	958	4.5

Al analizar la concentración diaria de PM10 (ver Fig.2) se evidencia un comportamiento similar al observado en las concentraciones diarias de PM2.5 dado que la estación de Carvajal reporta valores superiores a $50 \mu\text{g}/\text{m}^3$ mientras que las estaciones de Ministerio de Ambiente y el Tunal registran los valores más bajos (menores a $25 \mu\text{g}/\text{m}^3$). De igual manera las mayores concentraciones se reportan en el mes de febrero y diariamente entre las 6:00 y 10:00 a.m. La concentración diaria de PM10 para la ciudad oscila entre 26.7 y $44.4 \mu\text{g}/\text{m}^3$ con una media de $32.9 \mu\text{g}/\text{m}^3$.

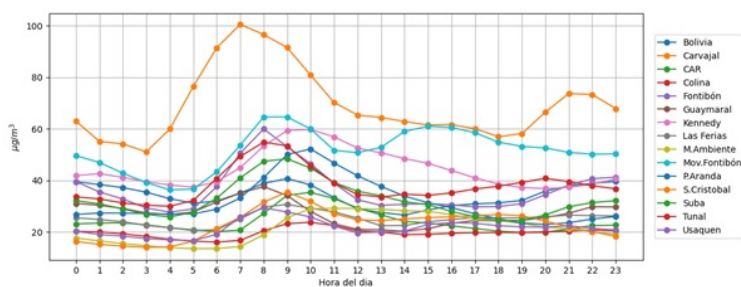


Figura 2: Concentración media de PM10. por cada hora del día (enero 2021 - junio 2023)

4.1. Modelo Extreme Gradient Boosting - XGB

Atendiendo a la presencia de valores no válidos en los registros de las estaciones estudiadas se reemplazaron los valores faltantes por el valor promedio por hora de todo el periodo. Posteriormente, se realizó el entrenamiento del modelo utilizando los datos de estación de Fontibón para PM10. El entrenamiento se realizó con los datos del 1 de enero de 2021 hasta el 31 de diciembre de 2022 y las pruebas se realizaron con los datos correspondientes a los registros del año 2023 (Fig.3).

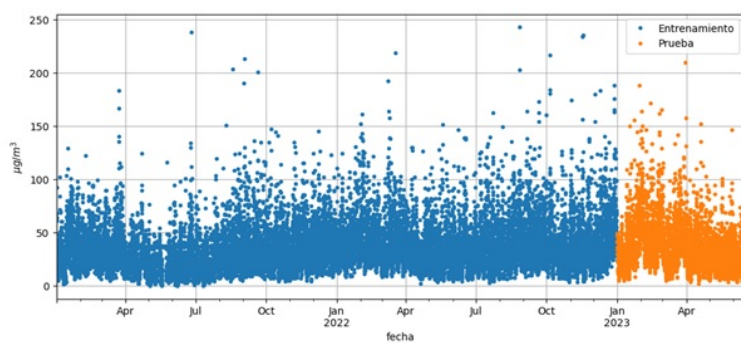


Figura 3: Conjunto de datos de entrenamiento y prueba del modelo.

Para el entrenamiento se utilizaron variables adicionales a las cuales se les asignó nivel de importancia



de mayor a menor así: registro de concentraciones de PM_{2.5} en la misma estación, hora del día, día del año, día de la semana y mes (Fig. 4).

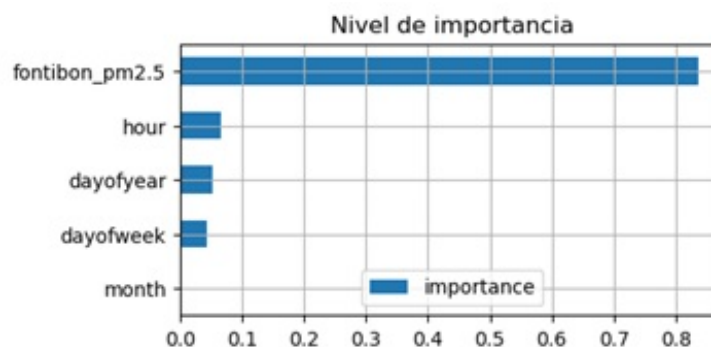


Figura 4: Nivel de importancia de otras variables utilizadas en el entrenamiento del modelo XGB

En la validación de la capacidad del modelo XGB para predecir las concentraciones de PM₁₀ se identificó que la predicción realizada se ajusta de forma consistente con los datos de prueba tal como se observa en la fig.5. Se obtuvo un puntaje de entrenamiento (training score) de 0.72 y un error de raíz cuadrada media (RSME) de 11.73.

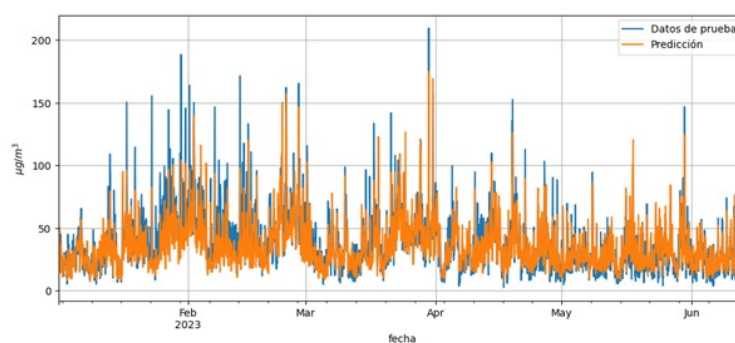


Figura 5: Validación de la predicción en el entrenamiento del modelo XGB

5. Discusión

El comportamiento de las concentraciones de PM_{2.5} y PM₁₀ registradas en la estación Carvajal han sido reportadas en estudios anteriores que señalan que pueden estar asociada con la ubicación geográfica de la estación cercana a avenidas con alto tráfico vehicular y el transporte regional de contaminantes desde los municipios vecinos al occidente de la ciudad en los cuales se realizan actividades industriales [26]. En contraste, las estaciones con menores concentraciones de PM_{2.5} y PM₁₀ se encuentran en áreas de la ciudad donde no se desarrollan actividades industriales o con presencia de zonas verdes (ej. Estación Centro de Alto Rendimiento – CAR, Tunal), los servicios ecosistémicos de los espacios verdes urbanos ha sido previamente reportados como favorecedores de la calidad de aire ya que las coberturas vegetales contribuyen a la captación de contaminantes presentes en el aire [27].

Por otra parte, el aumento de las concentraciones de PM_{2.5} y PM₁₀ observado en algunas franjas horarias coincide con las horas de mayor tráfico vehicular en la ciudad y con el inicio y el final de las jornadas laborales o actividades industriales [23]. Lo anterior puede ser de utilidad en el diseño e implementación de medidas preventivas y de control en la ciudad, tales como la entrada diferida de trabajadores, restricciones de movilidad (pico y placa) y áreas con restricciones de movilidad, entre otras. Se identificó que el



mes con mayores concentraciones de material particulado es febrero, de acuerdo con los reportes del RM-CAB [28] los incrementos pueden estar relacionados con contaminación derivada de incendios forestales en el oriente y sur del país; lo anterior teniendo en cuenta que febrero coincide la temporada seca.

Los valores medios de PM_{2.5} en algunas estaciones superan los niveles máximos permisibles diarios recomendados por la Resolución 2254 de 2017 que establece que los niveles máximos permisibles de PM_{2.5} para un tiempo de exposición 24 horas en Colombia son de $37 \mu\text{g}/\text{m}^3$ y $75 \mu\text{g}/\text{m}^3$ para el PM₁₀ [29]. Este aspecto es relevante teniendo en cuenta que la exposición a material particulado, especialmente el PM_{2.5}, está asociado con efectos negativos en la salud tales como enfermedades respiratorias y del sistema cardiovascular, entre otras [23].

Se identificaron estaciones con porcentajes de datos no válidos elevados, esto puede deberse a diferentes factores entre los cuales se encuentran fallas en el funcionamiento de los sensores y analizadores automáticos de las estaciones de monitoreo, realización de actividades de mantenimiento u otros eventos extraordinarios [28]. Por esta razón, para el entrenamiento y validación del modelo se seleccionaron los datos de la estación de monitoreo de Fontibón. Aunque se identificó que el modelo predice de forma acertada y consistente las concentraciones futuras de los contaminantes se considera necesario y un punto de buena práctica la prueba de otros modelos de predicción en las estaciones donde la calidad de los datos es deficiente.

Las predicciones realizadas por el modelo XGB se ajustan correctamente con los datos de prueba evidenciado en el puntaje de entrenamiento y el RMSE, de acuerdo con Ruiz Pastor [30] esto se debe a que el algoritmo de boosting utilizado por el modelo se enfoca en los ejemplos más difíciles de clasificar, esto debido a que durante el procesamiento secuencial en los árboles de decisión puede identificar las alertas, obteniendo así un mejor rendimiento.

En el entrenamiento del modelo XGB se utilizaron los datos de las concentraciones de PM_{2.5} en la misma estación y datos temporales; sin embargo, se recomienda la inclusión de información sobre los demás contaminantes (SO₂, NO₂, CO, O₃ y black carbon) y de las variables meteorológicas disponibles en el RMCAB tales como temperatura, radiación solar, humedad relativa, presión barométrica, precipitación, velocidad y dirección del viento. Lo anterior debido que estas variables son significativas y pueden proveer de mayor robustez y precisión a la predicción realizada [30], [31].

El modelo de pronóstico utilizado por Bogotá permite una predicción de 48 horas, el trabajo realizado por Castañeda et al. [31] demostró que la utilización de modelo XGB en combinación con modelo Light Gradient Boosting (LightGBM) aumenta la capacidad de predecir 72 horas a futuro la concentración de material particulado; sin embargo, debido a que la calidad y suficiencia de los datos en las diferentes estaciones varía es necesario que se implemente un modelo diferente para cada una de las estaciones y los contaminantes analizados.

La predicción de la calidad de aire realizada con el modelo XGB es complementaria a los modelos de pronóstico ya existentes en la ciudad y puede contribuir al fortalecimiento del sistema de alertas tempranas en algunas áreas de la ciudad, especialmente en aquellas con alto flujo vehicular o con presencia de zonas industriales. Adicionalmente, puede contribuir a la mejora en la aplicación de las medidas restrictivas que limitan la movilidad de diferentes vehículos automotores o que restringen la operación de fuentes fijas de emisión tales como zonas industriales.

6. Conclusiones

El modelo XGB permite la predicción acertada y robusta de las concentraciones de PM_{2.5} y PM₁₀ con lo cual se puede obtener información complementaria al sistema de pronóstico existente en la ciudad y contribuir al fortalecimiento de las acciones de monitoreo de la calidad de aire. Se recomienda que futuros trabajos incluyan información adicional sobre diferentes contaminantes y variables meteorológicas para incrementar la precisión de las predicciones; por otra parte, dada la divergencia de la información



proveniente de las diferentes estaciones de monitoreo de la ciudad se considera necesaria la aplicación de diferentes estrategias de predicción que se ajusten a la naturaleza de los datos disponibles.

Referencias

- [1] W. H. Organization, *Air pollution*, Online, 2023. dirección: https://www.who.int/health-topics/air-pollution#tab=tab_1.
- [2] U. EPA, *Particulate Matter (PM) Basics*, Online, 2022. dirección: <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>.
- [3] A. C. O'Regan y M. M. Nyhan, «Towards sustainable and net-zero cities: A review of environmental modelling and monitoring tools for optimizing emissions reduction strategies for improved air quality in urban areas,» *Environmental Research*, vol. 231, pág. 116-242, ago. de 2023, ISSN: 0013-9351. DOI: [10.1016/j.envres.2023.116242](https://doi.org/10.1016/j.envres.2023.116242). dirección: <http://dx.doi.org/10.1016/j.envres.2023.116242>.
- [4] M. Marć, M. Tobiszewski, B. Zabiegała, M. d. l. Guardia y J. Namieśnik, «Current air quality analytics and monitoring: A review,» *Analytica Chimica Acta*, vol. 853, págs. 116-126, ene. de 2015, ISSN: 0003-2670. DOI: [10.1016/j.aca.2014.10.018](https://doi.org/10.1016/j.aca.2014.10.018). dirección: <http://dx.doi.org/10.1016/j.aca.2014.10.018>.
- [5] R. Guo, Q. Zhang, X. Yu, Y. Qi y B. Zhao, «A deep spatio-temporal learning network for continuous citywide air quality forecast based on dense monitoring data,» *Journal of Cleaner Production*, vol. 414, pág. 137-568, ago. de 2023, ISSN: 0959-6526. DOI: [10.1016/j.jclepro.2023.137568](https://doi.org/10.1016/j.jclepro.2023.137568). dirección: <http://dx.doi.org/10.1016/j.jclepro.2023.137568>.
- [6] Y. Hu, X. Chen y H. Xia, «A hybrid prediction model of air quality for sparse station based on spatio-temporal feature extraction,» *Atmospheric Pollution Research*, vol. 14, n.º 6, pág. 101-765, jun. de 2023, ISSN: 1309-1042. DOI: [10.1016/j.apr.2023.101765](https://doi.org/10.1016/j.apr.2023.101765). dirección: <http://dx.doi.org/10.1016/j.apr.2023.101765>.
- [7] M. Kalhor y M. Bajoghli, «Comparison of AERMOD, ADMS and ISC3 for incomplete upper air meteorological data (case study: Steel plant),» *Atmospheric Pollution Research*, vol. 8, n.º 6, págs. 1203-1208, nov. de 2017, ISSN: 1309-1042. DOI: [10.1016/j.apr.2017.06.001](https://doi.org/10.1016/j.apr.2017.06.001). dirección: <http://dx.doi.org/10.1016/j.apr.2017.06.001>.
- [8] X. Yang, Q. Wu, R. Zhao et al., «New method for evaluating winter air quality: PM2.5 assessment using Community Multi-Scale Air Quality Modeling (CMAQ) in Xi'an,» *Atmospheric Environment*, vol. 211, págs. 18-28, ago. de 2019, ISSN: 1352-2310. DOI: [10.1016/j.atmosenv.2019.04.019](https://doi.org/10.1016/j.atmosenv.2019.04.019). dirección: <http://dx.doi.org/10.1016/j.atmosenv.2019.04.019>.
- [9] H.-M. Lee, R. J. Park, D. K. Henze et al., «<math>attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model,» *Environmental Pollution*, vol. 221, págs. 377-384, feb. de 2017, ISSN: 0269-7491. DOI: [10.1016/j.envpol.2016.11.088](https://doi.org/10.1016/j.envpol.2016.11.088). dirección: <http://dx.doi.org/10.1016/j.envpol.2016.11.088>.
- [10] S. Liu, S. Hua, K. Wang et al., «Spatial-temporal variation characteristics of air pollution in Henan of China: Localized emission inventory, WRF/Chem simulations and potential source contribution analysis,» *Science of The Total Environment*, vol. 624, págs. 396-406, mayo de 2018, ISSN: 0048-9697. DOI: [10.1016/j.scitotenv.2017.12.102](https://doi.org/10.1016/j.scitotenv.2017.12.102). dirección: <http://dx.doi.org/10.1016/j.scitotenv.2017.12.102>.
- [11] L. Johansson, A. Karppinen, M. Kurppa, A. Kousa, J. V. Niemi y J. Kukkonen, «An operational urban air quality model ENFUSER, based on dispersion modelling and data assimilation,» *Environmental Modelling and Software*, vol. 156, pág. 105-460, oct. de 2022, ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2022.105460](https://doi.org/10.1016/j.envsoft.2022.105460). dirección: <http://dx.doi.org/10.1016/j.envsoft.2022.105460>.
- [12] L. Zhang, J. Lin, R. Qiu et al., «Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model,» *Ecological Indicators*, vol. 95, págs. 702-710, dic. de 2018, ISSN: 1470-160X. DOI: [10.1016/j.ecolind.2018.08.032](https://doi.org/10.1016/j.ecolind.2018.08.032). dirección: <http://dx.doi.org/10.1016/j.ecolind.2018.08.032>.
- [13] L. Zhang, J. Lin, R. Qiu et al., «Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model,» *Ecological Indicators*, vol. 95, págs. 702-710, dic. de 2018, ISSN: 1470-160X. DOI: [10.1016/j.ecolind.2018.08.032](https://doi.org/10.1016/j.ecolind.2018.08.032). dirección: <http://dx.doi.org/10.1016/j.ecolind.2018.08.032>.



- [14] K. Das, N. Das Chatterjee, D. Jana y R. K. Bhattacharya, «Application of land-use regression model with regularization algorithm to assess PM2.5 and PM10 concentration and health risk in Kolkata Metropolitan,» *Urban Climate*, vol. 49, pág. 101 473, mayo de 2023, ISSN: 2212-0955. DOI: [10.1016/j.uclim.2023.101473](https://doi.org/10.1016/j.uclim.2023.101473). dirección: <http://dx.doi.org/10.1016/j.uclim.2023.101473>.
- [15] B. Bera, S. Bhattacharjee, N. Sengupta y S. Saha, «PM2.5 concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models,» *Environmental Challenges*, vol. 4, pág. 100 155, ago. de 2021, ISSN: 2667-0100. DOI: [10.1016/j.envc.2021.100155](https://doi.org/10.1016/j.envc.2021.100155). dirección: <http://dx.doi.org/10.1016/j.envc.2021.100155>.
- [16] C.-C. Liu, T.-C. Lin, K.-Y. Yuan y P.-T. Chiueh, «Spatio-temporal prediction and factor identification of urban air quality using support vector machine,» *Urban Climate*, vol. 41, pág. 101 055, ene. de 2022, ISSN: 2212-0955. DOI: [10.1016/j.uclim.2021.101055](https://doi.org/10.1016/j.uclim.2021.101055). dirección: <http://dx.doi.org/10.1016/j.uclim.2021.101055>.
- [17] S. M. Cabaneros y B. Hughes, «Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting,» *Environmental Modelling and Software*, vol. 158, pág. 105 529, dic. de 2022, ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2022.105529](https://doi.org/10.1016/j.envsoft.2022.105529). dirección: <http://dx.doi.org/10.1016/j.envsoft.2022.105529>.
- [18] B.-Y. Kim, Y.-K. Lim y J. W. Cha, «Short-term prediction of particulate matter (PM10 and PM2.5) in Seoul, South Korea using tree-based machine learning algorithms,» *Atmospheric Pollution Research*, vol. 13, n.º 10, pág. 101 547, oct. de 2022, ISSN: 1309-1042. DOI: [10.1016/j.apr.2022.101547](https://doi.org/10.1016/j.apr.2022.101547). dirección: <http://dx.doi.org/10.1016/j.apr.2022.101547>.
- [19] T. Chen y C. Guestrin, «XGBoost: A Scalable Tree Boosting System,» en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ép. KDD '16, ACM, ago. de 2016. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). dirección: <http://dx.doi.org/10.1145/2939672.2939785>.
- [20] K. Gui, H. Che, Z. Zeng et al., «Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model,» *Environment International*, vol. 141, pág. 105 801, ago. de 2020, ISSN: 0160-4120. DOI: [10.1016/j.envint.2020.105801](https://doi.org/10.1016/j.envint.2020.105801). dirección: <http://dx.doi.org/10.1016/j.envint.2020.105801>.
- [21] T. Thongthammachart, S. Araki, H. Shimadera, T. Matsuo y A. Kondo, «Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO2 and PM2.5 levels in Kansai region, Japan,» *Environmental Modelling and Software*, vol. 155, pág. 105 447, sep. de 2022, ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2022.105447](https://doi.org/10.1016/j.envsoft.2022.105447). dirección: <http://dx.doi.org/10.1016/j.envsoft.2022.105447>.
- [22] P.-Y. Wong, C.-Y. Hsu, J.-Y. Wu et al., «Incorporating land-use regression into machine learning algorithms in estimating the spatial-temporal variation of carbon monoxide in Taiwan,» *Environmental Modelling and Software*, vol. 139, pág. 104 996, mayo de 2021, ISSN: 1364-8152. DOI: [10.1016/j.envsoft.2021.104996](https://doi.org/10.1016/j.envsoft.2021.104996). dirección: <http://dx.doi.org/10.1016/j.envsoft.2021.104996>.
- [23] L. A. Rodríguez-Camargo, R. J. Sierra-Parada y L. C. Blanco-Becerra, «Análisis espacial de las concentraciones de PM2, 5 en Bogotá según los valores de las guías de la calidad del aire de la Organización Mundial de la Salud para enfermedades cardiopulmonares, 2014-2015,» *Biomédica*, vol. 40, n.º 1, págs. 137-152, mar. de 2020, ISSN: 0120-4157. DOI: [10.7705/biomedica.4719](https://doi.org/10.7705/biomedica.4719). dirección: <http://dx.doi.org/10.7705/biomedica.4719>.
- [24] S. D. de Ambiente, *Resumen ejecutivo del informe anual de calidad de aire*, Online, 2021. dirección: <http://rmcab.ambientebogota.gov.co/Pagesfiles/2022IE193351%20resumen%20informe%20anual%202021.pdf>.
- [25] S. D. de Ambiente, *Sobre el Modelo de Pronóstico de Calidad de Aire de Bogotá*, Online, 2023. dirección: <https://sdsgissaludbog.maps.arcgis.com/apps/MapSeries/index.html?appid=0a6ba64953d04fc592e9292b17c8bbdd>.
- [26] P. A. García Ávila y N. Y. Rojas, «Análisis del origen de PM10 y PM2.5 en Bogotá gráficos polares,» *Revista Mutis*, vol. 6, n.º 2, págs. 47-58, oct. de 2016, ISSN: 2256-1498. DOI: [10.21789/22561498.1150](https://doi.org/10.21789/22561498.1150). dirección: <http://dx.doi.org/10.21789/22561498.1150>.



- [27] R. Revelli y A. Porporato, «Ecohydrological model for the quantification of ecosystem services provided by urban street trees,» *Urban Ecosystems*, vol. 21, n.º 3, págs. 489-504, feb. de 2018, ISSN: 1573-1642. DOI: [10.1007/s11252-018-0741-2](https://doi.org/10.1007/s11252-018-0741-2). dirección: <http://dx.doi.org/10.1007/s11252-018-0741-2>.
- [28] R. de Monitoreo de Calidad de Aire de Bogotá - RMCAB, *Informe mensual de calidad de aire de Bogotá - febrero de 2022*, Online, 2022. dirección: <http://rmcab.ambientebogota.gov.co/Pagesfiles/informe%20mensual%20febrero%202022.pdf>.
- [29] M. de Ambiente y Desarrollo Sostenible, *Resolución 2254 de 2017 Por la cual se adopta la norma de calidad del aire ambiente y se dictan otras disposiciones*, Online, 2017. dirección: <https://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=82634>.
- [30] C. R. Pastor, *Técnicas de aprendizaje automático para el análisis de datos de calidad de aire*, Online, 2017. dirección: https://repositorio.uam.es/bitstream/handle/10486/679327/ruiz_pastor_carlos_tfg.pdf?sequence=1.
- [31] D. F. R. M. M. C. Castañeda Palencia J. Chavez Gonzalez, *Pronóstico de las PM2.5 y PM10 que incluye el impacto de las políticas de gestión ambiental en Bogotá*, Online, 2022. dirección: https://repository.javeriana.edu.co/bitstream/handle/10554/61509/attachment_1_213024-Casta%C3%B1eda-Chaves-Ram%C3%ADrez.pdf?sequence=1.

