

UNIVERSIDAD
NACIONAL
DE COLOMBIA

PROYECTO **CULTURAL, CIENTÍFICO Y COLECTIVO** DE NACIÓN

Evaluación de sensores de bajo costo para el monitoreo de PM_{2.5} en el Valle de Aburrá

Jenny Rocío Ríos Martínez

Yris Olaya Morales

Grupo de Investigación en Ciencias de la Decisión

Facultad de Minas, Universidad Nacional de Colombia
Medellín, Colombia

Universidad Nacional de Colombia

PROYECTO **CULTURAL, CIENTÍFICO Y COLECTIVO** DE NACIÓN



Sensores de bajo costo (SBC) de PM_{2.5}

En los últimos años ha habido un crecimiento constante en el uso de SBC para monitorear la calidad de aire, en especial en proyectos de ciudadanos científicos donde la comunidad tiene un papel importante no solo en la generación de nuevo conocimiento, sino en la participación ciudadana en la toma de decisiones (Mahajan et al., 2020).



Sensor Sensirion SPS30
CanAirIO



Sensor Plantower PMS7003
Unloquer



Dylos DC 1700-PM

Figura 1. Sensores de bajo costo (SBC) de PM_{2.5}

Desempeño de los SBC

El desempeño de los SBC se puede desarrollar evaluando diferentes características de los dispositivos como los son:

- ✓ Completitud de datos: se refiere a la cantidad de datos de los SBC que se obtuvieron con respecto a la cantidad esperada.
- ✓ Exactitud: evalúa la calidad de los datos al comparar las mediciones de los SBC con un valor de referencia.
- ✓ Precisión: se refiere a la consistencia de las lecturas de los SBC.
- ✓ Sesgo: error sistemático. Valor que se suma o resta a la medición verdadera

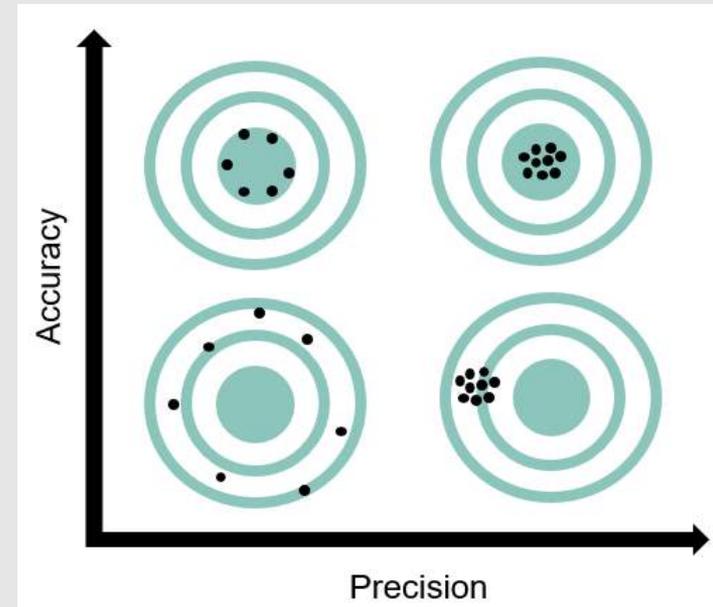


Figura 2. Precisión vs. Exactitud

Fuente: (Kang et al., 2022; Lewis et al., 2016; Williams et al., 2014).

Métricas de evaluación

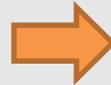
Para evaluar los las características de desempeño de los SBC, se pueden utilizar las siguientes métricas:

Coeficiente de correlación , r ,
para medir exactitud.



$$r = \frac{\sum_{i=1}^n \Delta c_{\text{estimated},i} \cdot \Delta c_{\text{true},i}}{\sqrt{\sum_{i=1}^n \Delta c_{\text{estimated},i}^2} \sqrt{\sum_{i=1}^n \Delta c_{\text{true},i}^2}}$$

Error absoluto medio (MAE),
Error cuadrático medio
(RMSE) para medir precisión.



$$MAE = \frac{\sum_{i=1}^n |c_{\text{estimated},i} - c_{\text{true},i}|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (c_{\text{estimated},i} - c_{\text{true},i})^2}{n}}$$

Sesgo medio normalizado
(MNB) para medir el sesgo o
bias.



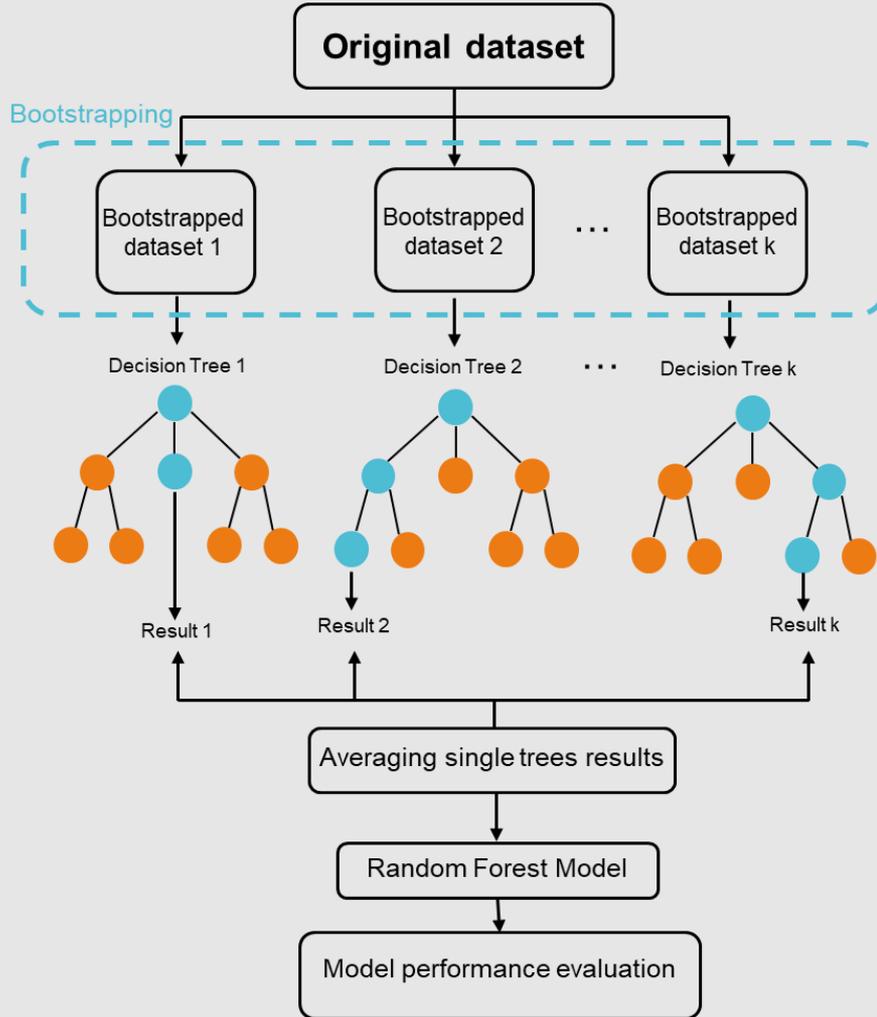
$$MNB = \frac{\sum_{i=1}^n (c_{\text{estimated},i} - c_{\text{true},i})}{\sum_{i=1}^n (c_{\text{true},i})}$$

Bosque Aleatorio o Random Forest (RF)

Un modelo RF es un algoritmo de aprendizaje automático para resolver problemas de regresión o clasificación (Breiman, 2001).

- **Principio básico:** Random forest es un algoritmo de aprendizaje de máquinas que usa árboles de decisión para modelar las relaciones entre variables.
- **Funcionamiento:** construcción de un conjunto de árboles de decisión utilizando un conjunto de datos de entrenamiento; el valor medio de ese conjunto de árboles de decisión se usa para predecir el valor de los resultados del modelo.

Bosque Aleatorio o Random Forest (RF)



1. Se tiene el dataset original con los datos de la colocación junto a un instrumento de referencia
2. Se divide el dataset en training y testing sets
3. Bootstrapping: hacer un re-sample aleatorio del training set con reemplazo. El tamaño del subset es el mismo del dataset original.
4. El nodo origen de cada árbol de decision se divide en sub-nodos, considerando un subset de las posibles variables explicatorias (m_{try}).
5. El resultado que se predice (concentración hora $PM_{2.5}$) está dado por el promedio del resultado de cada árbol de decision.
6. Se evalúa el desempeño del modelo para reproducir las observaciones. Se usan métricas de evaluación como RMSE y R^2 .

Figura 3. Pasos de Bosque aleatorio según (Zimmerman et al., 2018) y (Rodríguez-Galano et al., 2012)

Metodología

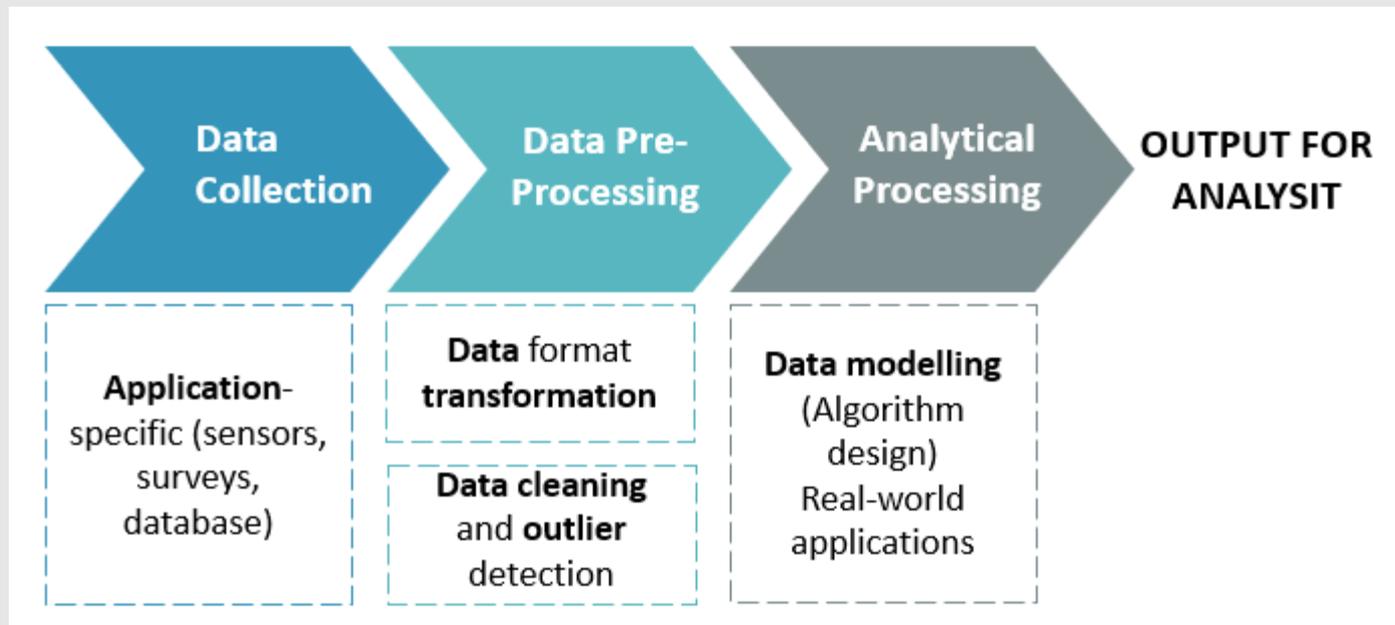


Figura 4. Pasos metodológicos propuestos en esta investigación

Recolección de datos

Área de estudio: Área Metropolitana del Valle de Aburrá. Municipio de Itagüí.

Contaminante medido: Material particulado $PM_{2.5}$

Tiempo de colocación de los dispositivos: 8 semanas, 18 mayo al 15 julio de 2021

Equipos usados:



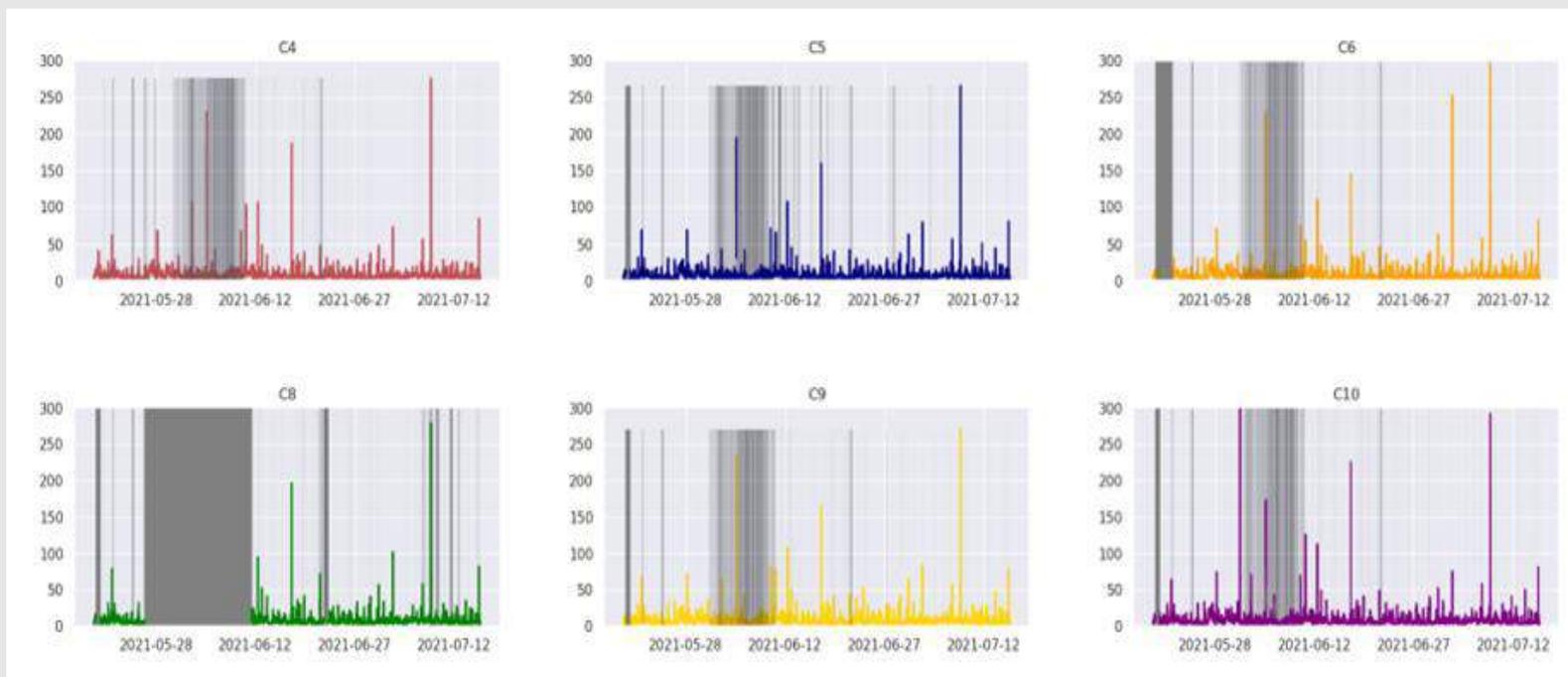
Figura 5. Estación de monitoreo SIATA I.E.
Concejo Municipal de Itagüí.
Dispositivo de referencia: BAM 1020
(*BAM-1020 Continuous Particulate Monitor, 2021*)



Figura 6. Sensores de bajo costo:
CanAirIO, Dylos y Unloquer

Pre-procesado de datos

1. Creación base datos unificada
2. Eliminación de valores perdidos para los SBC.
3. Eliminación de calidad de dato $\neq 1$ para los datos de referencia
4. Identificación y eliminación de outliers extremos ($>3IQR$) para los SBC.
5. Identificación de estadísticos



Resultados: Completitud de los datos

Sensor	Datos totales	Datos perdidos	% datos perdidos
Dylos 1	26378	57742	69%
Dylos 2	57009	27111	32%
Dylos 3	57019	27101	32%
CanAirIO 4	80816	3304	4%
CanAirIO 5	79566	4554	5%
CanAirIO 6	77693	6427	8%
CanAirIO 7	11641	72479	86%
CanAirIO 8	57947	26173	31%
CanAirIO 9	80371	3749	4%
CanAirIO 10	80446	3674	4%
Unloquer 1	63867	20253	24%
Unloquer 2	60771	23349	28%
Unloquer 3	64349	19771	24%
Unloquer 4	62171	21949	26%
Unloquer 5	66462	17658	21%
Unloquer 6	64238	19882	24%
Unloquer 7	62439	21681	26%
Unloquer 8	65112	19008	23%
Unloquer 9	21603	62517	74%
Unloquer 10	60610	23510	28%

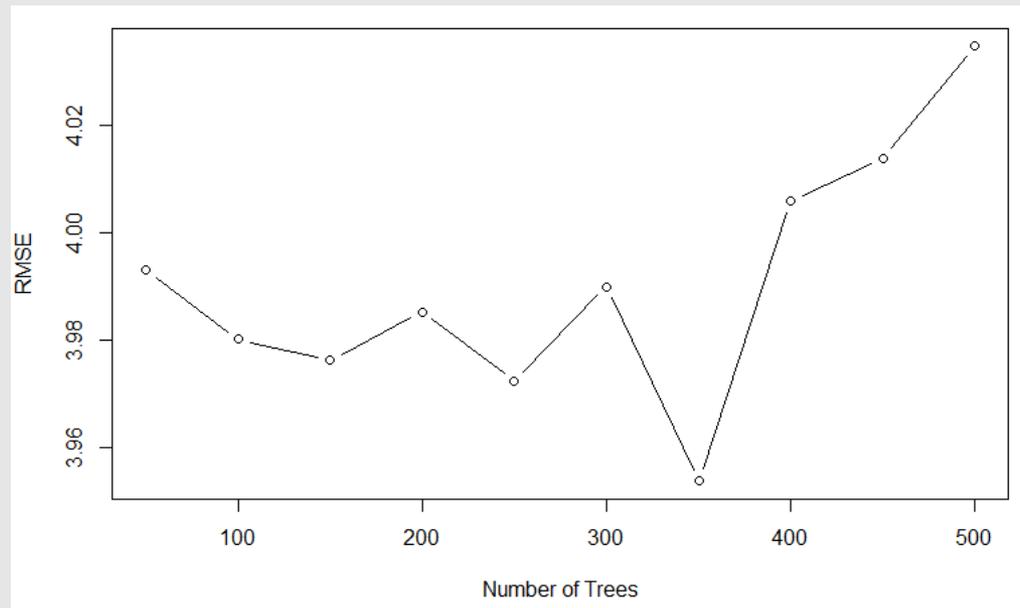
Se descartan para análisis posterior

Tabla 1. Datos totales y perdidos de los SBC

Resultados: Random Forest

Para cada SBC se buscó el número de árboles que permitiera obtener menor RMSE

Ejemplo: CanAirIO 4 – Sensirion SPS30



Resultados: Random Forest

Para cada SBC se buscó el número de árboles que permitiera obtener menor RMSE

	Regresión lineal simple			Random Forest		
	r	RMSE	MAE	r	RMSE	MAE
C4	0.84	3.997	3.216	0.89	3.951	3.205
C5	0.82	4.103	3.313	0.86	4.011	3.302
C6	0.80	4.243	3.441	0.83	4.135	3.431
C9	0.82	4.104	3.318	0.85	4.016	3.307
C10	0.83	4.049	3.260	0.88	3.982	3.249

Conclusiones

- Las limitaciones de los SBC en términos de exactitud y precisión pueden afectar su capacidad para reflejar el estado real de la calidad del aire en un lugar determinado. Por tanto no se deben usar como dispositivos de tipo regulatorio.
- La pérdida de datos se puede dar por problemas de suministro de energía, problemas en el envío y/o almacenamiento de datos en los SBC.
- Los SBC se utilizan en proyectos de ciencia ciudadana para dar información de tipo no regulatoria de a la comunidad y asistir en el planteamiento de políticas para reducir exposición personal a la contaminación del aire.
- Los modelos de regresión lineal y RF pueden ser usados para conocer y mejorar el desempeño de los SBC.
- Revisión del desempeño de los SBC periódico para validar su correcto funcionamiento.

Referencias

- *BAM-1020 Continuous Particulate Monitor*. (2021). Met One Instruments. <https://metone.com/products/bam-1020/>
- CanAirIO. (2021). *CanAirIO*. Citizen network for monitoring air quality. <https://canair.io/index.html>
- *DC1700-PM PM2.5/PM10 AQM*. (2022). <http://www.dylosproducts.com/dcpmaq.html>
- Dylos Corporation. (2022, octubre 10). *Dylos Corporation. Air quality monitoring innovation*. <http://www.dylosproducts.com/>
- Kang, Y., Aye, L., Ngo, T. D., & Zhou, J. (2022). Performance evaluation of low-cost air quality sensors: A review. *Science of The Total Environment*, 818, 151769. <https://doi.org/10.1016/j.scitotenv.2021.151769>
- Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., Smith, K. R., Buckley, J. W., Ellis, M., Gillot, S. R., & White, A. (2016). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions*, 189, 85-103. <https://doi.org/10.1039/C5FD00201J>
- Mahajan, S., Chung, M.-K., Martinez, J., Olaya, Y., Helbing, D., & Chen, L.-J. (2022). Translating citizen-generated air quality data into evidence for shaping policy. *Humanities and Social Sciences Communications*, 9(1), Art. 1. <https://doi.org/10.1057/s41599-022-01135-2>
- Sensirion. (2021). *SPS30—PM2.5 Sensor for HVAC and air quality applications SPS30*. <https://sensirion.com/products/catalog/SPS30/>
- Williams, R., Kilaru, V., Snyder, E., Kaufman, A., Dye, T., Rutter, A., Russell, A., & Hafner, H. (2014a). *Air Sensor Guidebook*. <https://www.epa.gov/air-sensor-toolbox/how-use-air-sensors-air-sensor-guidebook>
- Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Haurlyiuk, A., Robinson, E. S., Robinson, A. L., & R. Subramanian. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques*, 11(1), 291–313. <https://doi.org/10.5194/amt-11-291-2018>

Gracias

Universidad Nacional de Colombia

PROYECTO **CULTURAL, CIENTÍFICO Y COLECTIVO** DE NACIÓN