

PREDICCIÓN DE LAS CONCENTRACIONES DE MATERIAL PARTICULADO EN LA CIUDAD DE SANTIAGO DE CALI APLICANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL

Ing. Hiran Jeison Flórez Pinzón
PhD. Juan Pablo Silva Vinasco



Facultad de
Ingeniería

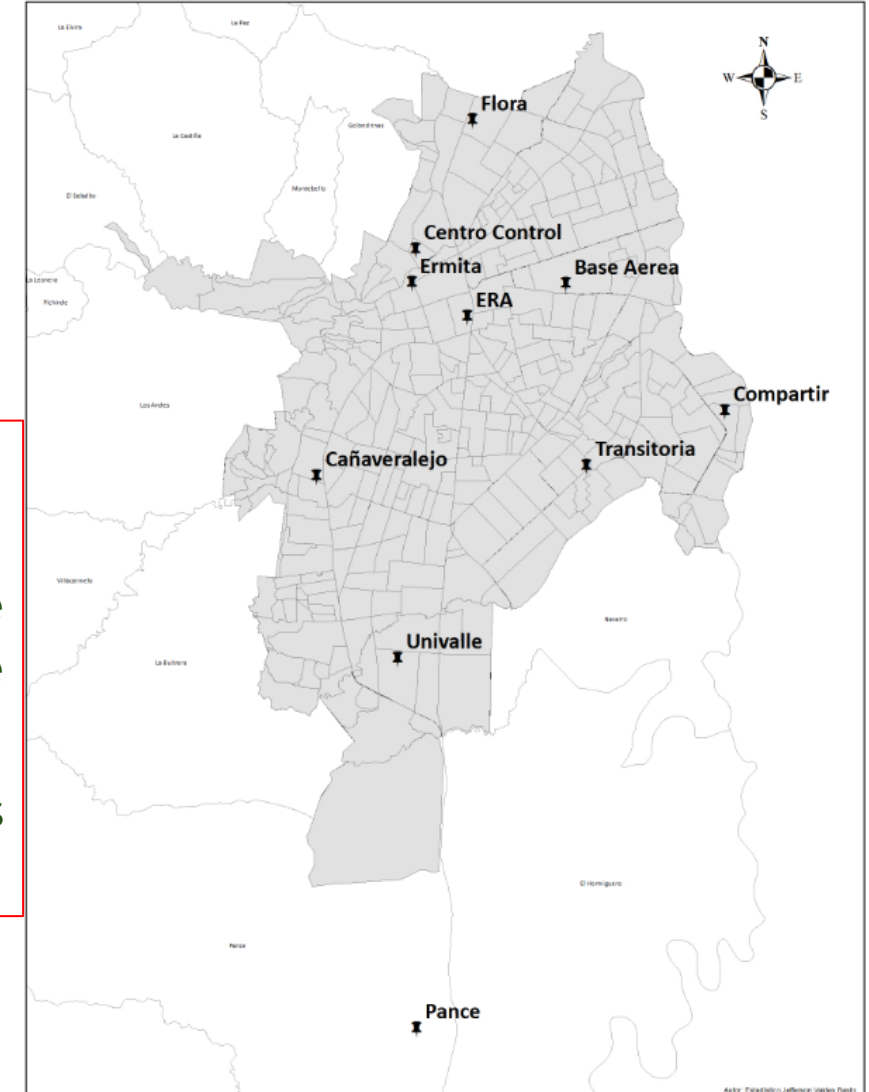
Escuela de Ingeniería de Recursos
Naturales y del Ambiente

Contexto del problema



La contaminación atmosférica debida al material particulado (PM_{10} y $PM_{2,5}$) representa un riesgo para la salud pública

- No hay continuidad en la medición
- La ciudad carece de una herramienta automática de pronóstico para la toma de decisiones preventivas
- Necesidad de anticiparse a los posibles eventos de contaminación



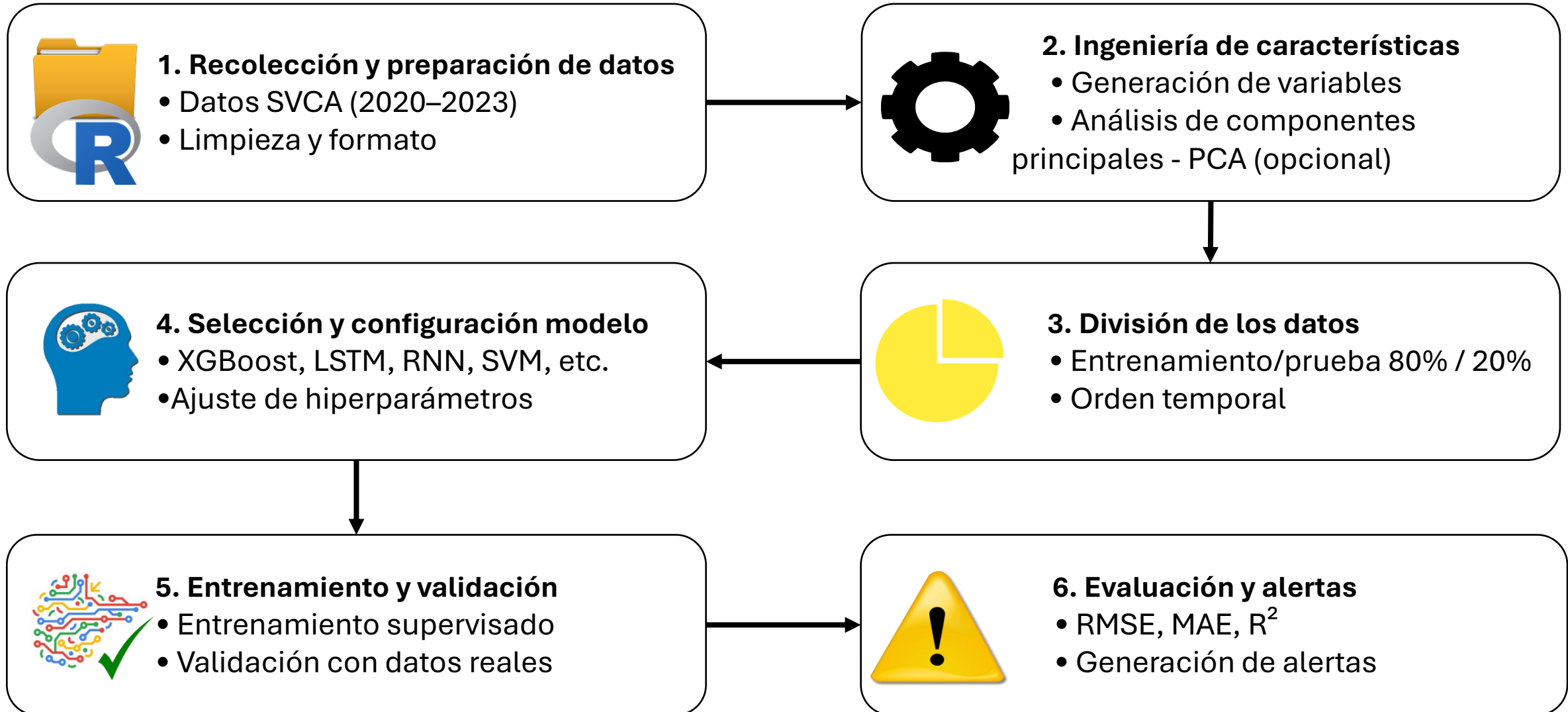
DAGMA (2024)



Inteligencia Artificial – IA

Herramienta para la predicción de las concentraciones de material particulado

Metodología

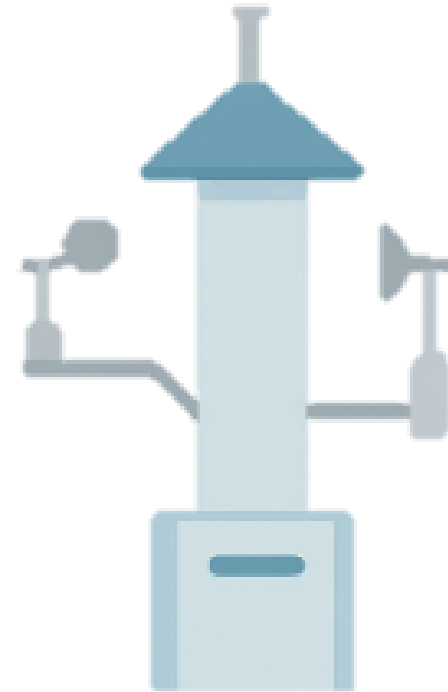


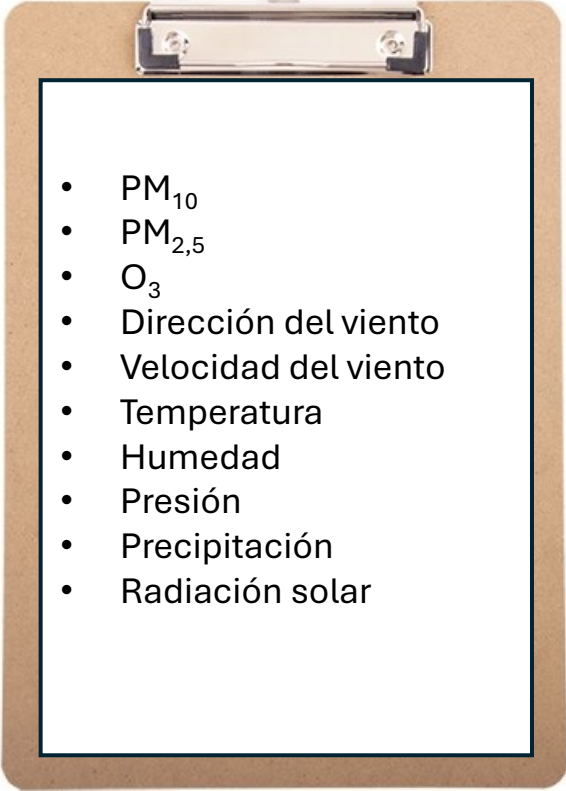
Metodología



Estaciones seleccionadas: Compartir y Universidad del Valle (Univalle)

Datos horarios obtenidos del Subsistema de Información sobre la Calidad del Aire – SISAIRE y de la Autoridad Ambiental (DAGMA)



- 
- PM₁₀
 - PM_{2,5}
 - O₃
 - Dirección del viento
 - Velocidad del viento
 - Temperatura
 - Humedad
 - Presión
 - Precipitación
 - Radiación solar



87600 datos por año deberían ser medidos

Se pierden entre 13000 y 20000 datos por año, es decir, entre el 15% y el 23% del total

Metodología

Modelo	Ventajas	Desventajas	Mejor Caso de Uso
Red Neuronal	- Captura relaciones no lineales- Flexible en arquitectura	- Entrenamiento lento- Puede sobreajustarse	Cuando hay muchos datos y relaciones complejas
SVM – Máquina de vectores de soporte	- Precisión alta en espacios no lineales- Bueno con pocos datos	- Costoso computacionalmente- Difícil de ajustar	Series con patrones claros y estables
XGBoost – Boosting de gradiente	- Rápido y eficiente- Importancia de variables interpretable	- Puede perder rendimiento con ruido- Sensible a datos atípicos	Datos limpios y bien estructurados
LSTM – Memoria a largo y corto plazo	- Ideal para datos secuenciales- Capta dependencias a largo plazo	- Necesita muchos datos- Alto costo de entrenamiento	Series temporales con rezagos importantes

Resultados

ESTACIÓN UNIVALLE

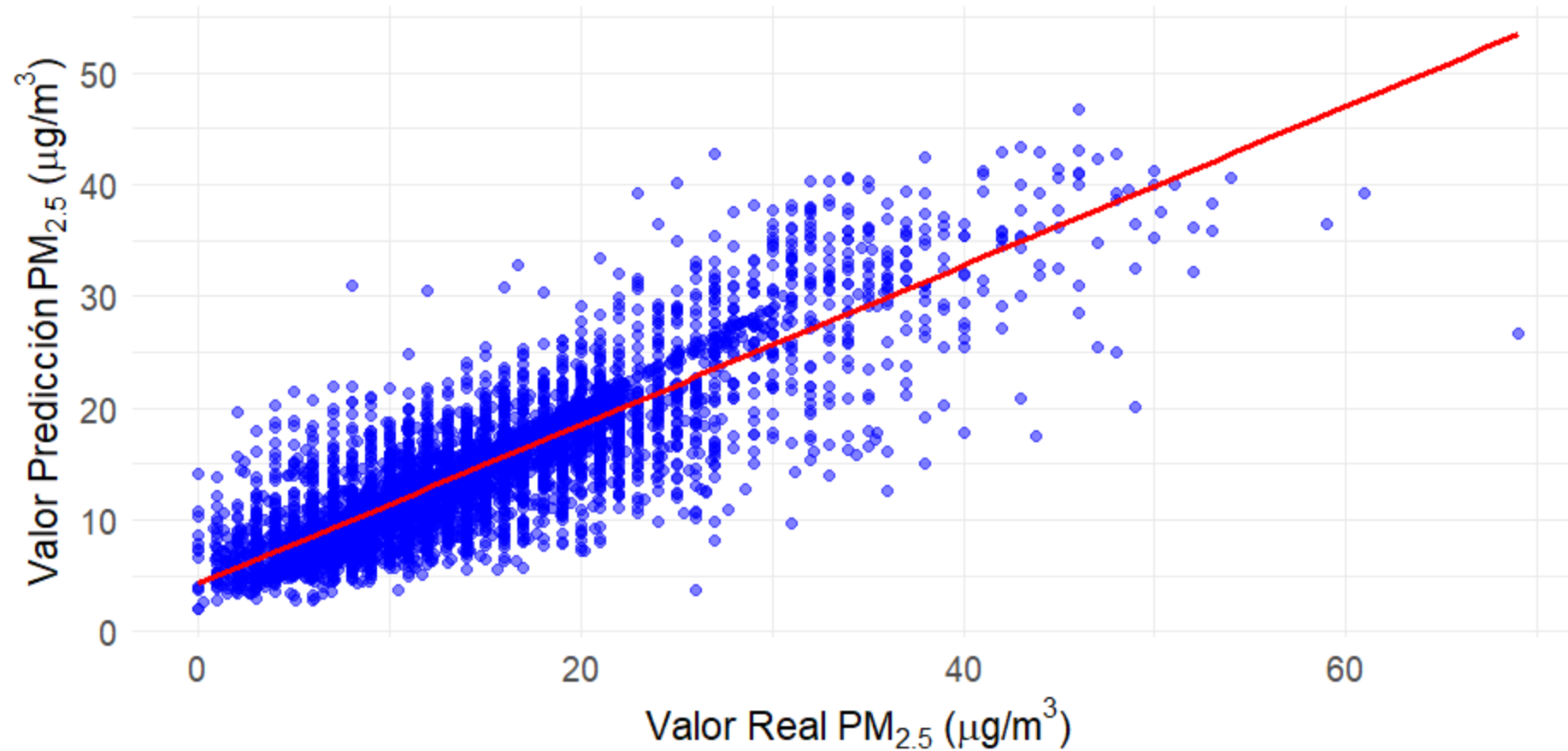
Contaminante	Modelo	RMSE	MAE	R ²
PM ₁₀	XGBoost	5,68	3,02	0,76
PM ₁₀	LSTM	4,41	2,70	0,71
PM _{2,5}	XGBoost	4,03	2,58	0,75
PM _{2,5}	LSTM	3,45	1,98	0,67

ESTACIÓN COMPARTIR

Contaminante	Modelo	RMSE	MAE	R ²
PM ₁₀	XGBoost	15,14	9,05	0,57
PM ₁₀	LSTM	13,23	7,01	0,62
PM _{2,5}	XGBoost	6,32	4,16	0,31
PM _{2,5}	LSTM	5,77	3,36	0,43

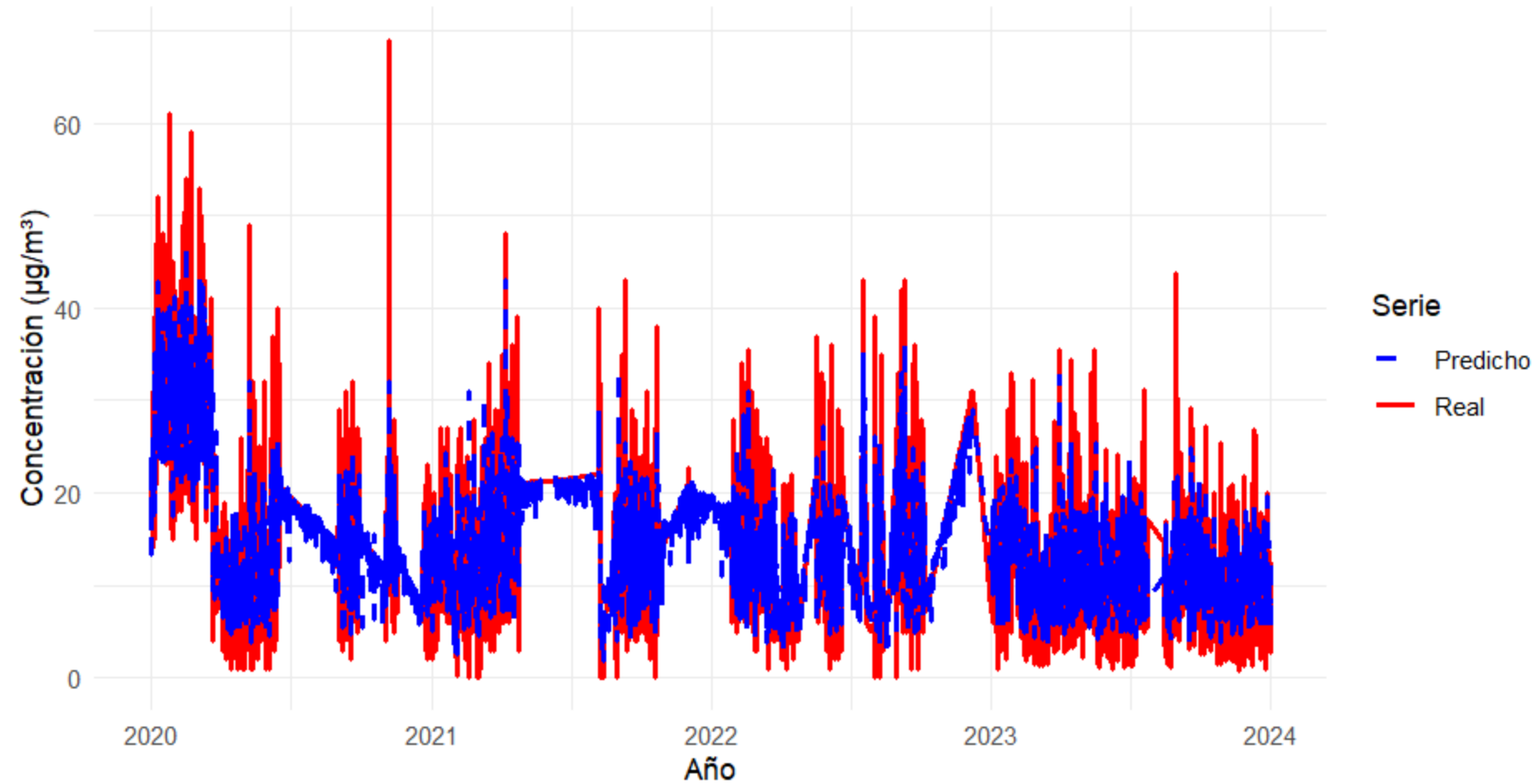
Resultados

Univalle - LSTM Real vs Predicción



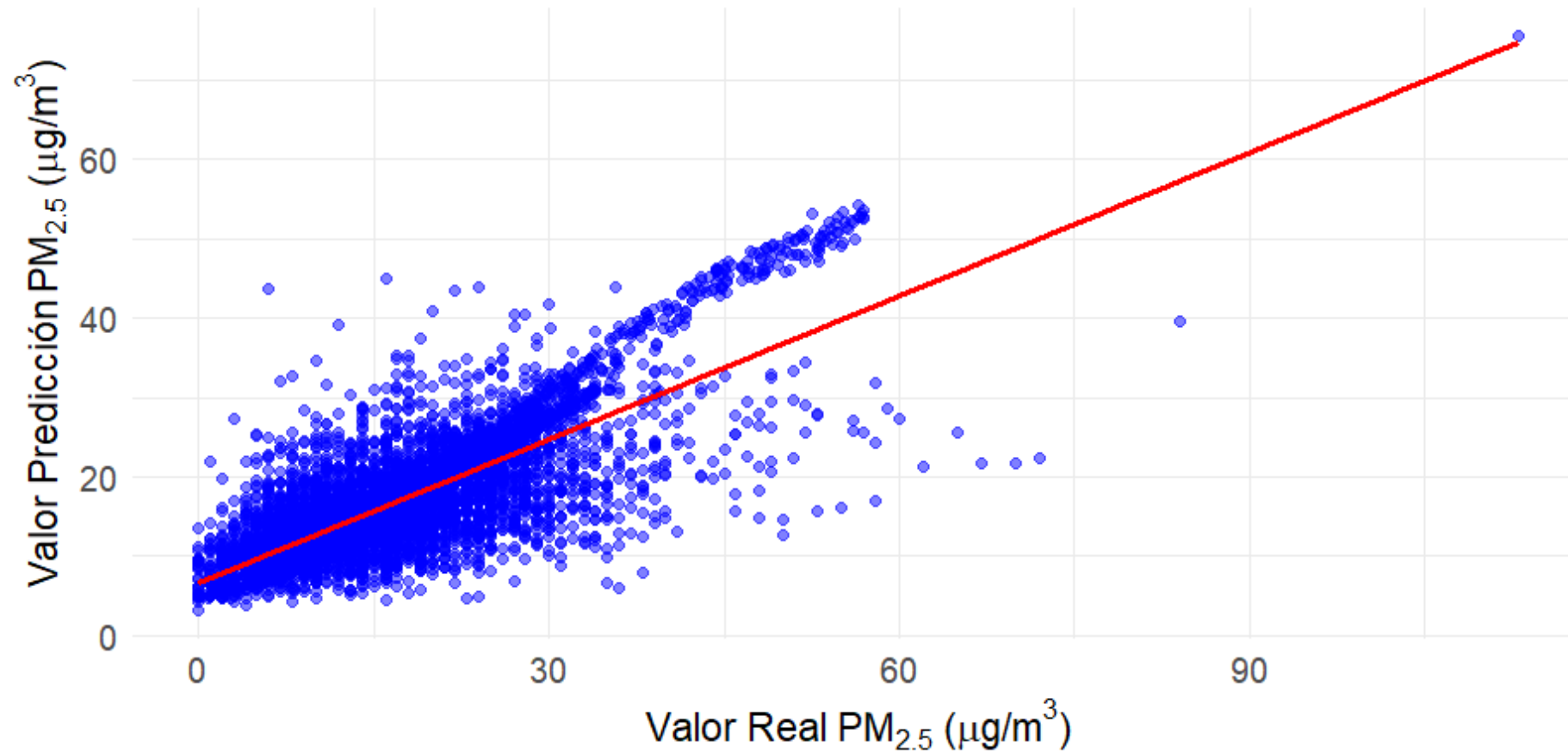
Resultados

Univalle Serie temporal PM_{2.5}: Real vs Predicción con LSTM



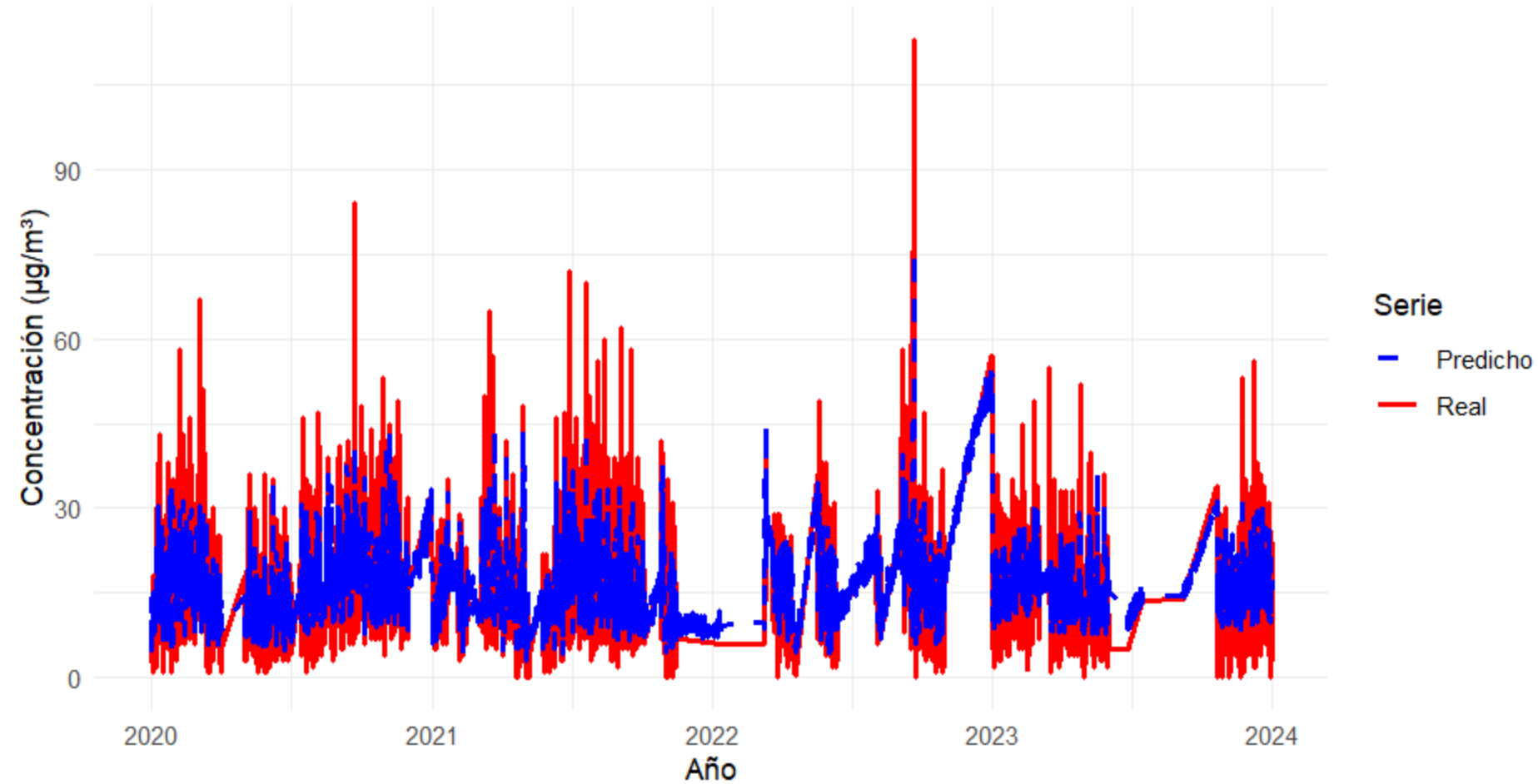
Resultados

Compartir - LSTM Real vs Predicción



Resultados

Compartir Serie temporal PM_{2.5}: Real vs Predicción con LSTM



Resultados

Sistema de Alerta Temprana

XGBoost

Contaminante	Estación	Umbral µg/m ³	Valor medido µg/m ³	Valor predicción µg/m ³	Mensaje
PM ₁₀	Univalle	75	39,70	43,99	Sin alerta: Predicción dentro del rango aceptable
PM ₁₀	Compartir	75	53,24	55,87	Sin alerta: Predicción dentro del rango aceptable
PM _{2,5}	Univalle	37	5,60	5,96	Sin alerta: Predicción dentro del rango aceptable
PM _{2,5}	Compartir	37	16,00	13,99	Sin alerta: Predicción dentro del rango aceptable

Resultados

Sistema de Alerta Temprana

LSTM

Contaminante	Estación	Umbral $\mu\text{g}/\text{m}^3$	Valor medido $\mu\text{g}/\text{m}^3$	Valor predicción $\mu\text{g}/\text{m}^3$	Mensaje
PM ₁₀	Univalle	75	39,70	36,34	Sin alerta: Predicción dentro del rango aceptable
PM ₁₀	Compartir	75	53,24	53,20	Sin alerta: Predicción dentro del rango aceptable
PM _{2,5}	Univalle	37	5,60	5,47	Sin alerta: Predicción dentro del rango aceptable
PM _{2,5}	Compartir	37	16,00	14,72	Sin alerta: Predicción dentro del rango aceptable

Conclusiones

Entre el 15% y el 23% de los datos horarios anuales de cada estación no están disponibles, lo cual representa un reto para la modelación precisa y continua de la calidad del aire.

La pérdida sistemática de datos implica la necesidad de aplicar técnicas robustas de limpieza e imputación para asegurar predicciones confiables.

Se resalta la importancia de implementar herramientas automáticas de pronóstico como apoyo en la toma de decisiones preventivas en salud pública.

Las estaciones Compartir y Univalle presentan comportamiento suficiente para el desarrollo y validación de modelos predictivos basados en inteligencia artificial.



**Facultad de
Ingeniería**

**Escuela de Ingeniería de Recursos
Naturales y del Ambiente**

Contacto:

hiran.florez@correounivalle.edu.co

juan.silva@correounivalle.edu.co

Más información



<https://casap.science/>



casap@casap.science