

# **Aplicación de series de tiempo para el análisis predictivo de la calidad del aire en Bogotá**

Paola Andrea Garcia Altamirano

Andrés Felipe Ariza Pardo

Asesor

Fernando Luis Carrascal Porras

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

## Nota de Aceptación

---

Director: Luis Ángel Anillo Arrieta

---

Jurado 1

---

Jurado 2

## Resumen

Este proyecto aplica modelos de series de tiempo para predecir la calidad del aire en Bogotá utilizando datos públicos. Se busca identificar patrones y tendencias que permitan anticipar variaciones en la contaminación, proporcionando indicadores clave para la toma de decisiones en salud pública. Además de reforzar los conocimientos adquiridos en la especialización, el proyecto contribuye al conocimiento formal de la comunidad y promueve el uso de herramientas de análisis predictivo en problemáticas ambientales.

***Palabras clave:*** Modelo, Calidad, Aire, Bogotá, Series.

### **Abstract**

This project applies time-series models to predict air quality in Bogotá using public data. It seeks to identify patterns and trends that allow for anticipating variations in pollution, providing key indicators for public health decision-making. In addition to reinforcing the knowledge acquired in the specialization, the project contributes to the community's formal knowledge base and promotes the use of predictive analytics tools for environmental issues.

***Keywords:*** Model, Quality, Air, Bogotá, Series.

## Tabla de Contenido

Introducción .....	9
Justificación .....	10
Objetivos.....	11
Objetivo General .....	11
Objetivos Específicos.....	11
Marco Referencial.....	12
Metodología .....	15
Recolección y Preparación de Datos.....	15
Normalización.....	15
Transformación A Formato Supervisado .....	16
División de los Datos .....	16
Modelado .....	17
Modelo ARIMAX (ARIMA sin Variables Exógenas).....	17
Modelo Lstm (Long Short-Term Memory).....	18
Modelo Nn (Red Neuronal).....	18
Modelo Prophet .....	19
Modelo R3 (Random Forest Recursive Forecasting).....	20
Predicción y Evaluación.....	21
Resumen de Herramientas Utilizadas .....	22
Resultados .....	23
Recomendaciones .....	25
Conclusiones.....	27

Referencias..... 29

## Lista de Figuras

<b>Figura 1</b> <i>Gráfico Pronostico ARIMA</i> .....	17
<b>Figura 2</b> <i>Gráfico pronostico LSTM</i> .....	18
<b>Figura 3</b> <i>Gráfico Pronostico Red Neuronal Simple</i> .....	19
<b>Figura 4</b> <i>Gráfico Pronostico Prophet</i> .....	20
<b>Figura 5</b> <i>Gráfico Random Forest RF</i> .....	21

**Lista de Tablas**

<b>Tabla 1</b> <i>Resultados Aplicación de Modelos</i> .....	23
--	----

## Introducción

En una ciudad como Bogotá, donde viven más de siete millones de personas, la calidad del aire se ha convertido en una preocupación constante. Día a día, muchos ciudadanos respiran aire con niveles de contaminación que superan los límites recomendados por la Organización Mundial de la Salud. Esto no solo afecta al medio ambiente, sino también a la salud de miles de personas, especialmente a quienes sufren de afecciones respiratorias o cardiovasculares.

Este problema no es nuevo, pero sigue siendo urgente. Por eso, contar con herramientas que nos permitan anticipar los momentos y lugares en los que la contaminación será más crítica puede marcar una gran diferencia, por eso buscamos desarrollar un modelo predictivo que nos ayude a entender cómo evoluciona la calidad del aire en Bogotá. Para lograrlo, se utilizarán datos históricos, información meteorológica y técnicas modernas de análisis, como el aprendizaje automático y las series de tiempo.

Con esta herramienta, será posible anticipar episodios de alta contaminación y actuar con mayor eficacia. Las autoridades podrán tomar decisiones informadas y emitir alertas para que la ciudadanía pueda adaptar sus actividades para reducir su exposición y proteger su salud.

## Justificación

Por esta razón es importante contar con herramientas que permitan predecir la calidad del aire. Los modelos basados en series de tiempo hacen posible anticipar cambios en los niveles de contaminación y entregar información útil para diseñar medidas de prevención y generar alertas a la población.

Este proyecto busca crear un modelo predictivo que use técnicas de aprendizaje automático y análisis de series de tiempo para pronosticar la calidad del aire en Bogotá. Con él se pueden lograr varios beneficios. La población recibirá alertas tempranas que le permitan tomar decisiones y reducir la exposición a contaminantes. Las autoridades podrán anticipar episodios de contaminación y aplicar medidas preventivas como restricciones vehiculares o advertencias sanitarias.

Además, contar con predicciones más precisas ayudará a mejorar las políticas ambientales y a diseñar estrategias más efectivas para reducir emisiones en sectores como el transporte y la industria. La combinación de datos históricos con información meteorológica y de sensores en tiempo real aumentará la precisión del modelo y permitirá adaptarlo a cambios estacionales o situaciones inesperadas.

En resumen, este proyecto promueve el uso de inteligencia artificial para mejorar la gestión de la calidad del aire y reducir el impacto de la contaminación en la salud y el bienestar de los habitantes de Bogotá.

## **Objetivos**

### **Objetivo General**

Desarrollar un sistema de pronóstico de la calidad del aire en Bogotá mediante modelos estadísticos y de series de tiempo aplicados a datos históricos de concentraciones de contaminantes atmosféricos (PM10) estación de monitoreo de Fontibón, con el fin de anticipar niveles críticos de contaminación y apoyar la toma de decisiones en salud y gestión ambiental.

### **Objetivos Específicos**

Preprocesar y consolidar los datos históricos de calidad del aire provenientes de diferentes mediciones del componente PM10, organizándolos en una estructura unificada que permita su análisis conjunto por estación y contaminante.

Aplicar modelos de series de tiempo para generar pronósticos individuales en alguna estación de monitoreo de Bogotá.

Comparar el rendimiento de los modelos predictivos, evaluando su precisión y capacidad de generalización en la predicción de concentraciones futuras de contaminantes atmosféricos.

## Marco Referencial

La calidad del aire en ciudades como Bogotá representa un desafío creciente para la salud pública, especialmente para grupos vulnerables como niños, adolescentes y adultos mayores. La exposición prolongada a contaminantes atmosféricos, en particular al material particulado (PM10), se asocia con un incremento en enfermedades respiratorias, cardiovasculares y neurológicas. Investigaciones respaldadas por la Organización Mundial de la Salud han demostrado que estos efectos son más severos durante etapas sensibles del desarrollo, como la infancia y el embarazo.

Entre 2018 y 2022, localidades como Bosa, Kennedy y Ciudad Bolívar registraron las tasas más altas de mortalidad por infecciones respiratorias agudas, lo que refleja la estrecha relación entre la contaminación atmosférica y el deterioro de la salud en zonas densamente pobladas. Ante este panorama, se hace evidente la necesidad de contar con herramientas tecnológicas que permitan no solo monitorear la calidad del aire, sino también anticipar escenarios críticos y diseñar estrategias de intervención oportunas.

Según la Resolución 2254 de 2017, en Colombia el nivel máximo permisible de PM10 es de  $50 \mu\text{g}/\text{m}^3$  como promedio anual y  $100 \mu\text{g}/\text{m}^3$  en un periodo de 24 horas, reduciéndose a  $75 \mu\text{g}/\text{m}^3$  para 24 horas desde julio de 2018. De forma progresiva, para el año 2030 se establecerá un límite anual de  $30 \mu\text{g}/\text{m}^3$ . Estos valores reflejan una política orientada a cumplir estándares más estrictos y a reducir los riesgos asociados a la exposición prolongada a partículas contaminantes. Las proyecciones indican que, si se mantienen las acciones de control de emisiones, la modernización del transporte y la adopción de tecnologías limpias en sectores productivos, Colombia podría lograr mejoras significativas en su calidad del aire.

En este contexto, el Machine Learning se ha consolidado como una herramienta clave para abordar problemas complejos relacionados con el pronóstico de la calidad del aire. Esta rama de la inteligencia artificial permite desarrollar modelos capaces de aprender patrones a partir de datos históricos y generar predicciones precisas. Su integración con plataformas de Big Data y sistemas en la nube facilita el análisis de grandes volúmenes de información en tiempo real, lo que resulta esencial para la gestión ambiental urbana.

Bogotá cuenta con el sistema IBOCA (Índice Bogotano de Calidad del Aire y Riesgo en Salud), administrado por la Secretaría Distrital de Ambiente, que ofrece datos abiertos y actualizados diariamente sobre contaminantes como el PM10. En este proyecto se emplearán los registros de la estación de monitoreo de Fontibón, lo que permitirá un análisis contextualizado y basado en evidencia. El acceso a esta información no solo fomenta la transparencia institucional, sino que también incentiva la participación ciudadana y el desarrollo de soluciones tecnológicas en salud ambiental.

Para el análisis y modelado de los datos, se utilizará Python, un lenguaje de programación ampliamente adoptado por su versatilidad y por contar con bibliotecas especializadas como Scikit-learn, Skforecast y Prophet. Estas herramientas permiten implementar, comparar y evaluar distintos algoritmos de predicción, facilitando la selección del modelo más eficiente. Entrenados con datos históricos de IBOCA, estos modelos pueden identificar relaciones no evidentes entre variables ambientales y generar pronósticos que ayuden a anticipar situaciones de riesgo para la salud pública.

Por eso, la combinación de técnicas de Machine Learning con datos ambientales locales constituye una estrategia sólida para fortalecer la toma de decisiones en tiempo real. Esto no solo incrementa la capacidad de respuesta institucional ante episodios de alta contaminación, sino que

también contribuye a garantizar el derecho constitucional de los colombianos a gozar de un ambiente sano, consagrado en los artículos 79 y 80 de la Constitución Política de Colombia.

## **Metodología**

El objetivo de esta investigación fue predecir la concentración de PM10 (partículas en suspensión atmosférica menores a 10 micrómetros) utilizando modelos de series temporales. Para este estudio, no se consideraron variables exógenas. Asimismo, se aplicaron y compararon diferentes modelos con el fin de identificar aquel con menor error, ya sea en términos de MAE o RMSE. A continuación, se describe el procedimiento utilizado para el análisis:

### **Recolección y Preparación de Datos**

Se utilizó una serie temporal univariada correspondiente a los registros de concentración de PM10 para la zona de Fontibón. Aunque se disponía de más información de esta y otras zonas, así como de otras variables de calidad del aire, se tomó esta serie como referencia. El intervalo de tiempo seleccionado fue de frecuencia horaria. El conjunto de datos contiene dos columnas principales: la fecha (eje X) y los valores de PM10 (eje Y). Las etapas realizadas fueron:

- Lectura de los datos desde un archivo .xlsx.
- Filtro por localidad (Fontibón).
- Conversión de la columna de fecha al formato de fecha y hora.
- Ordenamiento de los registros por fecha.
- Establecimiento del índice temporal (datetime) para facilitar la manipulación.
- Manejo de datos faltantes mediante dropna.

### **Normalización**

Dado que los modelos de redes neuronales son sensibles al rango de los datos, se normalizó la variable PM10 utilizando MinMaxScaler, transformando los valores a un rango entre 0 y 1.

### ***Transformación A Formato Supervisado***

Para entrenar los modelos de aprendizaje supervisado, se transformó la serie temporal en un conjunto de datos tipo ventana móvil (sliding window) utilizando retardos (lags). Se generaron secuencias de longitud fija (24 horas anteriores) como entrada, con el fin de predecir el valor siguiente.

### ***División de los Datos***

Se separaron las últimas 24 horas del conjunto de datos como conjunto de prueba, utilizando el resto de la serie para entrenamiento. La predicción se enfocó en replicar el comportamiento de esas últimas 24 horas conocidas.

## Modelado

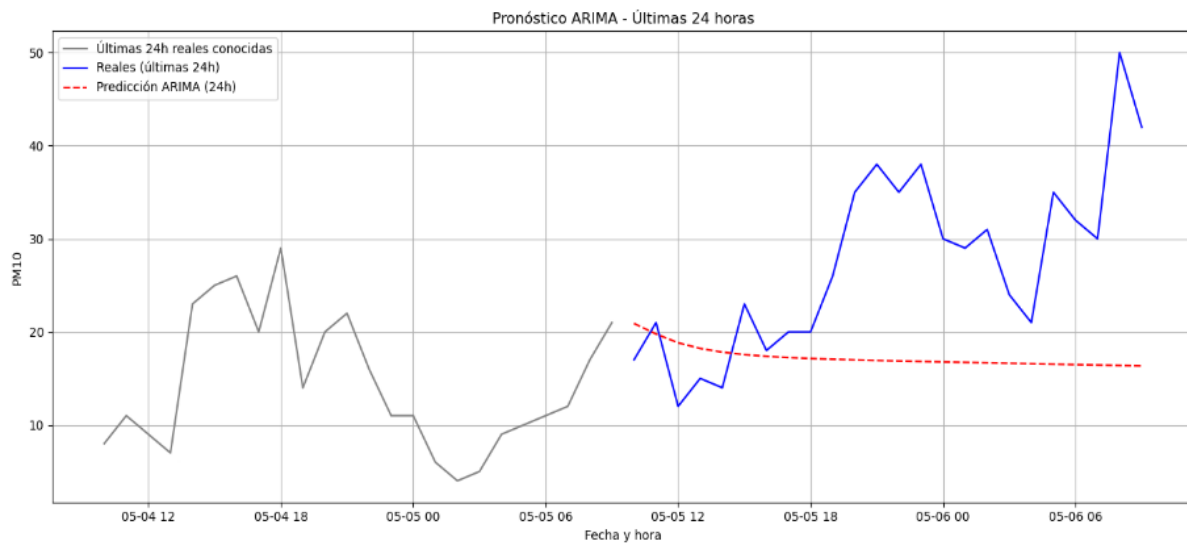
### *Modelo ARIMAX (ARIMA sin Variables Exógenas)*

Este modelo utiliza un enfoque estadístico para pronosticar series de tiempo. Se empleó la función `auto_arima` para encontrar automáticamente los mejores parámetros (p, d, q) del modelo ARIMA. El mejor modelo encontrado fue un ARIMA (2,0,3) con un intercepto y las métricas de evaluación fueron:

- MAE (Error Absoluto Medio): 11.52
- RMSE (Raíz del Error Cuadrático Medio): 14.31

## Figura 1

### *Gráfico Pronóstico ARIMA*



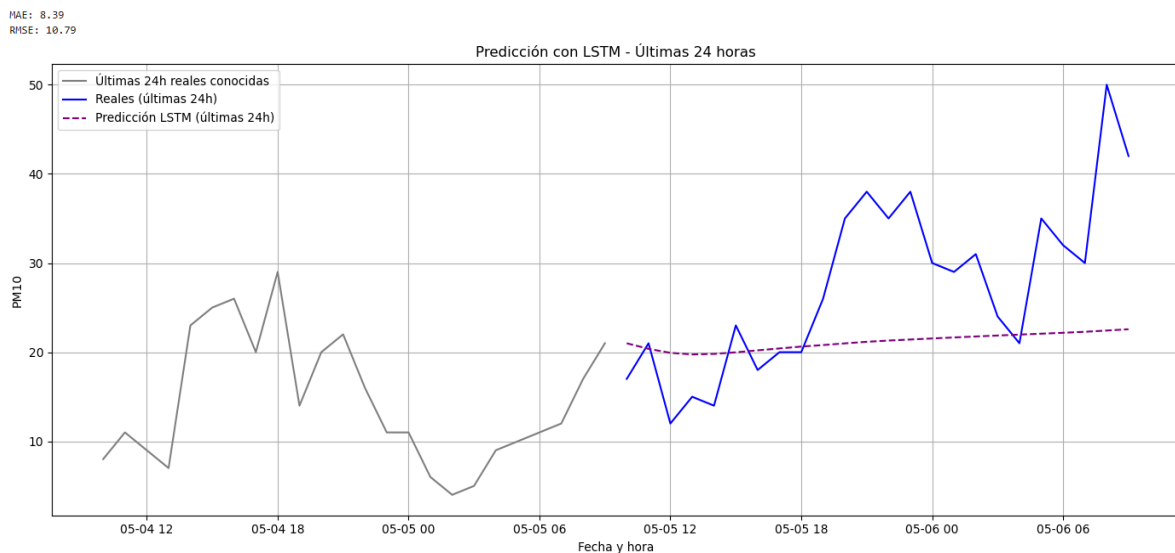
### ***Modelo Lstm (Long Short-Term Memory)***

Este modelo emplea un tipo de red neuronal recurrente (RNN) diseñado para manejar datos secuenciales y capturar dependencias a largo plazo. Los datos fueron escalados utilizando MinMaxScaler y transformados en secuencias de las 24 horas previas para predecir el siguiente valor. El modelo LSTM consta de 50 unidades, una capa de salida densa; las métricas de evaluación fueron:

- RMSE (Raíz del Error Cuadrático Medio): 10.79
- MAE (Error Absoluto Medio): 8.39

### **Figura 2**

#### *Gráfico pronostico LSTM*



### ***Modelo Nn (Red Neuronal)***

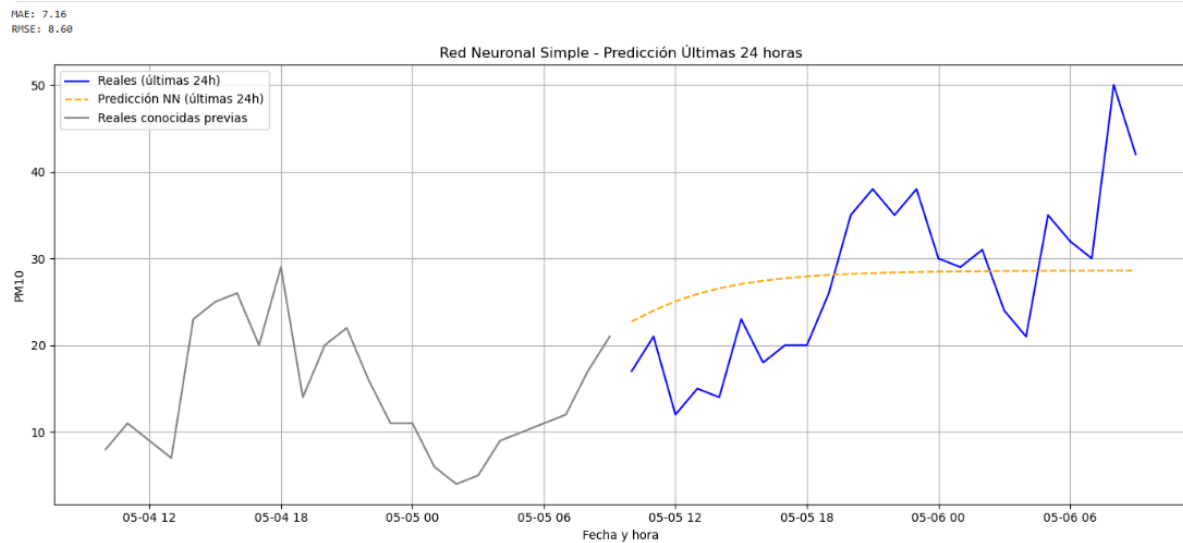
Este modelo utiliza una red neuronal secuencial para predecir valores basados en observaciones rezagadas. Los datos fueron escalados, y se crearon dos características rezagadas (lag\_1 y lag\_2) a partir de la variable objetivo. La red neuronal se compone de tres capas densas

(32, 16 y 1 neurona, respectivamente) con activación ReLU para las capas ocultas. Las métricas de evaluación fueron:

- MAE (Error Absoluto Medio): 7.16
- RMSE (Raíz del Error Cuadrático Medio): 8.6

### Figura 3

*Gráfico Pronostico Red Neuronal Simple*



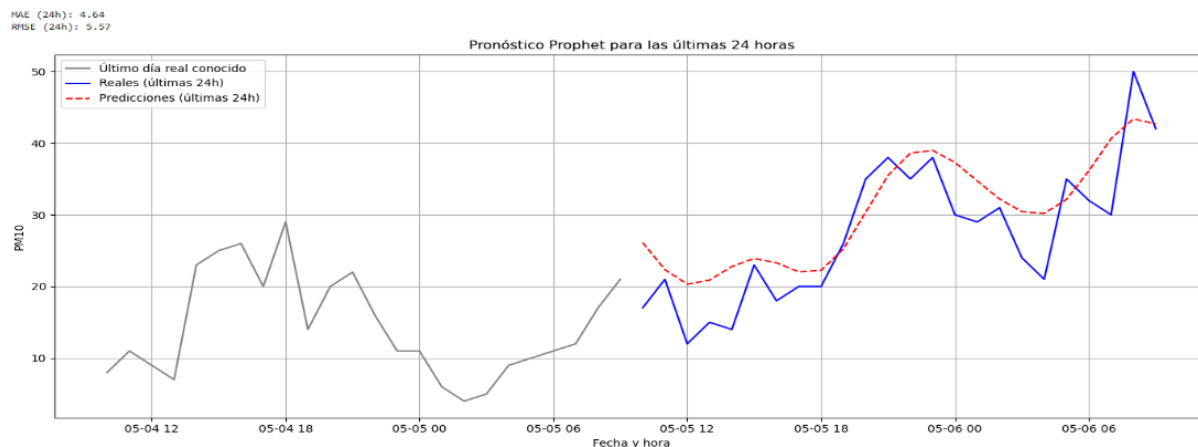
### *Modelo Prophet*

Prophet es un modelo de pronóstico desarrollado por Facebook, diseñado para datos de series de tiempo que exhiben fuertes efectos estacionales y días festivos. El modelo fue entrenado con datos históricos y luego utilizado para predecir las siguientes 24 horas con las siguientes métricas de evaluación:

- MAE (Error Absoluto Medio): 4.64
- RMSE (Raíz del Error Cuadrático Medio): 5.57

## Figura 4

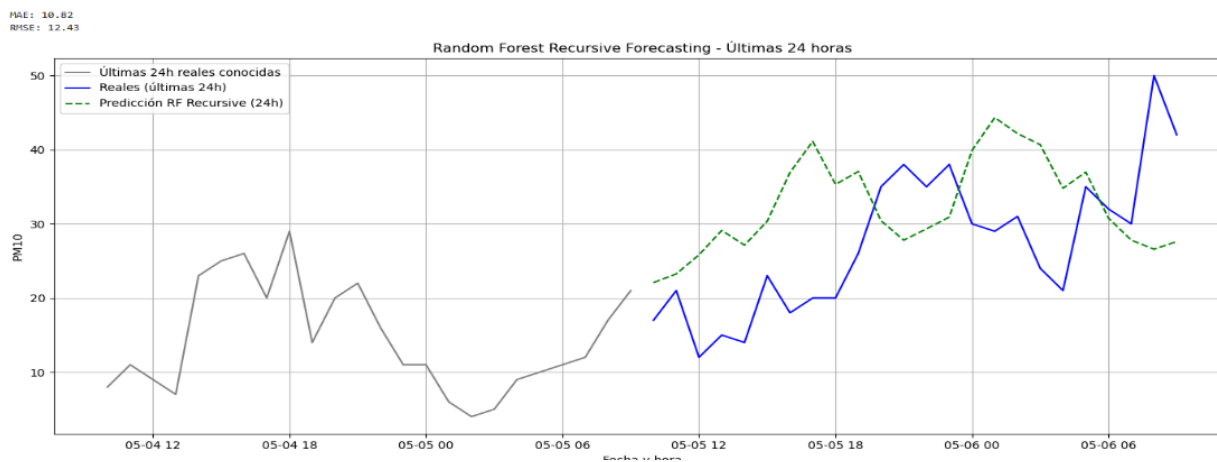
### Gráfico Pronóstico Prophet



### Modelo R3 (Random Forest Recursive Forecasting)

Este modelo emplea un Regresor de Bosque Aleatorio (Random Forest Regressor) en una configuración de pronóstico recursivo. Esto significa que las predicciones para futuros pasos de tiempo se basan en valores previamente predichos. Se utilizaron características rezagadas (`lag_1`, `lag_2`, `lag_3`, `lag_4`, `lag_5`, `lag_6`, `lag_12`, `lag_24`) como entradas, junto con la hora y el día de la semana. Las métricas de evaluación fueron:

- MAE (Error Absoluto Medio): 10.82
- RMSE (Raíz del Error Cuadrático Medio): 12.43

**Figura 5***Gráfico Random Forest RF***Predicción y Evaluación**

Las últimas 24 horas fueron predichas con cada modelo y se compararon los resultados con los valores reales, utilizando las siguientes métricas:

- **MAE (Error Absoluto Medio):** Error Medio Absoluto es un indicador que mide el promedio de la diferencia absoluta entre las predicciones y los valores reales. Esta penaliza todos los errores por igual y es fácil de interpretar, si el MAE da un valor de 5 quiere decir que nos equivocamos en la predicción 5 unidades.

- **RMSE (Raíz del Error Cuadrático Medio) o Error de Raíz Cuadrada Media** es el promedio de los errores al cuadrado. Y se diferencia del MAE por que penaliza más fuerte los grandes errores debido a que los eleva al cuadrado. Al igual que el MAE es fácil de interpretar e indica las unidades las cuales nos equivocamos al predecir

Además, se visualizaron las predicciones frente a los valores reales, permitiendo evaluar la capacidad de los modelos para capturar patrones locales y variaciones características de la serie.

## Resumen de Herramientas Utilizadas

- Librerías: pandas, numpy, scikit-learn, matplotlib, tensorflow.keras, pmdarima, fbprophet.

- Lenguaje: Python 3.12.4
- Frecuencia de la serie: Horaria

## Resultados

**Tabla 1**

*Resultados Aplicación de Modelos*

Modelo	Descripción breve	MAE (Error RMSE (Raíz Error Medio)		Observaciones
		Absoluto	Cuadrático	
ARIMA (2,0,3)	Modelo estadístico clásico sin variables exógenas.	11.52	14.31	Adecuado para capturar tendencias simples, pero menos preciso frente a métodos de ML.
LSTM	Red neuronal recurrente especializada en secuencias y dependencias a largo plazo.	8.39	10.79	Mejor rendimiento que ARIMA; captura patrones temporales más complejos.
Red Neuronal (NN)	Red secuencial con tres capas densas (32-16-1 neuronas).	7.16	8.60	Desempeño superior a LSTM y ARIMA; buena para series con patrones no lineales.
Prophet	Modelo de Facebook para series con estacionalidad y festivos.	4.64	5.57	Mejor desempeño general; más robusto a valores atípicos y estacionalidad.
Random Forest (R3)	Pronóstico recursivo con múltiples retardos y variables de tiempo.	10.82	12.43	Rendimiento intermedio; útil si se incorporan más variables predictoras.

Prophet es el modelo con mejor desempeño, teniendo los valores de error más bajos (MAE = 4.64 y RMSE = 5.57). lo que indica que sus predicciones son cercanas a los valores reales con desviaciones mínimas, Su capacidad para manejar estacionalidad y datos con valores atípicos lo hace especialmente adecuado para la serie de PM10 en Bogotá, la Red Neuronal también es un buen modelo ya que ocupó el segundo lugar en precisión (MAE = 7.16, RMSE = 8.60), mostrando un buen ajuste en los patrones lineales, el LSTM presentó un rendimiento aceptable (MAE = 8.39, RMSE = 10.79), pero no es lo suficientemente bueno a comparación de los modelos anteriores, el 3. Random Forest (R3) y ARIMA (2,0,3) son los modelos con menor precisión relativa. Aunque ARIMA es el más sencillo y útil para series lineales, queda rezagado frente a los enfoques de aprendizaje automático.

En general el análisis confirma que, para la predicción de PM10 en la estación Fontibón, Prophet es la mejor opción, seguido de la Red Neuronal. Los modelos estadísticos clásicos y Random Forest, sin variables adicionales, ofrecen menor exactitud y serían más útiles si se enriquecen con variables meteorológicas y datos de otras estaciones.

## Recomendaciones

Con base en los resultados obtenidos y el proceso de modelado predictivo desarrollado, se proponen las siguientes recomendaciones:

**Ampliar el análisis con variables meteorológicas:** Para futuros trabajos, se sugiere incorporar variables climáticas como temperatura, humedad relativa, velocidad y dirección del viento, ya que estos factores tienen un efecto significativo sobre la dispersión y concentración de contaminantes atmosféricos. Su inclusión podría mejorar la precisión de los modelos y enriquecer el análisis multivariable.

Basándonos en los resultados de evaluación presentados para los diferentes modelos de pronóstico, se recomienda el Modelo Prophet para la tarea de anticipar las variaciones en los niveles de contaminación del aire en Bogotá.

El Modelo Prophet, desarrollado por Facebook, ha demostrado consistentemente ser el más preciso en las pruebas realizadas. Esto se evidencia en sus métricas de evaluación significativamente inferiores (lo cual es mejor) en comparación con los otros modelos:

- MAE (Error Absoluto Medio): 4.64
- RMSE (Raíz del Error Cuadrático Medio): 5.57

Un MAE y RMSE más bajos indican que las predicciones del Modelo Prophet se desvían, en promedio, menos de los valores reales, lo que se traduce en una mayor exactitud y fiabilidad para la toma de decisiones.

Además de su precisión superior, el Modelo Prophet es particularmente adecuado para series de tiempo que presentan estacionalidad y efectos de días festivos, características comunes en los datos de calidad del aire. Su diseño intuitivo y su capacidad para manejar datos con valores atípicos lo convierten en una herramienta robusta y práctica para este tipo de aplicación.

Diseñar un sistema de alerta temprana: A partir de los resultados obtenidos, es posible desarrollar una herramienta automatizada de alerta que notifique a las autoridades y a la ciudadanía cuando se pronostiquen niveles críticos de contaminación. Esta aplicación puede integrarse en plataformas de monitoreo ambiental ya existentes.

Extender el análisis a más estaciones y periodos de tiempo: Se sugiere expandir el alcance del estudio incluyendo datos de años anteriores y de estaciones adicionales para robustecer los modelos y asegurar una representación más amplia del comportamiento de la calidad del aire en la ciudad.

Fortalecer la accesibilidad y actualización de los datos: Para facilitar futuros trabajos, se recomienda a las entidades encargadas del monitoreo ambiental garantizar el acceso oportuno y estandarizado a los datos, así como su documentación técnica, lo cual favorecerá su aprovechamiento en investigaciones académicas y decisiones institucionales.

## Conclusiones

El proceso de ETL fue un paso esencial, a pesar de que los datos son públicos y están disponibles en cualquier momento, la calidad no es siempre la mejor, ya que puede que falten datos por diferentes factores y la transformación de estos es esencial para que los modelos reciban una estructura uniforme con intervalos bien definidos.

Se compararon modelos clásicos y modelos de machine Learning, los modelos de machine Learning tienen mayor flexibilidad y capacidad para modelar las relaciones no lineales, también exigen una mayor cantidad de datos tanto intrínsecos de la predicción como valores exógenos los cuales no se incluyeron en el análisis con el fin de que todos los modelos estuvieran en las mismas condiciones.

En el modelo de prophet se identificó una gran facilidad de implementación y de esta forma destacó sobre los otros modelos, es de pronto con más pruebas, datos y un análisis más exhaustivo, otros modelos puedan superar los resultados del prophet debido a las limitaciones que posee.

El enfoque recursivo fue clave para mejorar la precisión de todos los modelos, ya que sin esta técnica los valores de error aumentaban de forma consistente. Esto teniendo en cuenta que el enfoque recursivo introduce un pequeño error acumulativo adicional de forma inevitable y con el aumento de las iteraciones u horizonte de predicción.

Existen otros modelos que no se evaluaron como los de LSTM bidireccionales o Transformers para series temporales, integración de variables exógenas y técnicas de multi-step forecasting que pueden cambiar los resultados encontrados en estos modelos.

En general el MAE en los modelos fue menor al RMSE, esto muestra la no linealidad de los datos y sus predicciones, o visto desde otro enfoque, el RMSE penalizó los datos atípicos que encontró y esto generó mayores valores de error en esta medición

En ese sentido se puede decir que, en el periodo analizado, los niveles de PM10 en Bogotá se mantuvieron por debajo del límite máximo permisible establecido por la Resolución 2254 de 2017 ( $100 \mu\text{g}/\text{m}^3$  en 24 horas), con valores que alcanzaron picos cercanos a  $50 \mu\text{g}/\text{m}^3$ . No obstante, la tendencia ascendente observada en las últimas horas indica un incremento en la concentración de material particulado que, de persistir, podría acercarse a niveles de riesgo moderado según el Índice de Calidad del Aire. El modelo Prophet, con un bajo margen de error (MAE = 4.64, RMSE = 5.57), permitió anticipar este comportamiento, lo que evidencia su utilidad para el monitoreo y la prevención de episodios críticos de contaminación.

## Referencias

- Alonso, S. (2021). Estudio y predicciones de calidad del aire en la ciudad de Madrid [Trabajo de fin de grado, UNED]. Repositorio e-Espacio UNED.  
<https://oai.espacio.uned.es/server/api/core/bitstreams/ae232d5d-89d2-47ec-ae74-48d97a1d63f1/content>
- Bai, S., Zico Kolter, J., & Koltun, V. (2017). Modeling long- and short-term temporal patterns with deep neural networks. arXiv. <https://doi.org/10.48550/arXiv.1703.07015>
- Dutta, S., & Basak, H. (2023). Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations. *Atmospheric Environment*, 297, 119987. <https://doi.org/10.1016/j.atmosenv.2023.119987>
- El País. (2024, 19 de abril). IBOCA: cómo funciona la alerta por mala calidad del aire en Bogotá y cuándo se activa cada fase. <https://elpais.com/america-colombia/20240419/iboca-como-funciona-la-alerta-por-mala-calidad-del-aire-en-bogota-y-cuando-seactivacada-fase.html>
- Kumar, A., & Rai, H. (2022). Time series forecasting (TSF) using various deep learning models. arXiv. <https://doi.org/10.48550/arXiv.2204.11115>
- Mehta, A., & Kulkarni, P. (2021). Comparing Prophet and Deep Learning to ARIMA in forecasting wholesale food prices. arXiv. <https://doi.org/10.48550/arXiv.2107.12770>
- Ministerio de Ambiente y Desarrollo Sostenible. (2017, 1 de noviembre). Resolución 2254 de 2017: Por la cual se adopta la norma de calidad del aire ambiente y se dictan otras disposiciones [Resolución]. <https://www.minambiente.gov.co/wp-content/uploads/2021/10/Resolucion-2254-de-2017.pdf>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Education.

- Observatorio Ambiental. (2019). Estudio de la calidad del aire en la Universidad Sergio Arboleda en jornada del día sin moto y sin carro: Análisis de la concentración de PM2.5. Universidad Sergio Arboleda.  
[https://www.usergioarboleda.edu.co/centros\\_e\\_institutos/observatorioambiental/opinion/estudio-calidad-aire-dia-sin-carro-2019](https://www.usergioarboleda.edu.co/centros_e_institutos/observatorioambiental/opinion/estudio-calidad-aire-dia-sin-carro-2019)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peña, A. (2020). Pronóstico y prevención de la contaminación por PM10 en la red de monitoreo de calidad del aire de Bogotá (RMCAB) [Tesis de maestría, Universidad de los Andes]. Repositorio Institucional Uniandes.  
<https://repositorio.uniandes.edu.co/server/api/core/bitstreams/3213778f-12c4-42d0-a6f1-382066753dae/content>
- Rincón, L. (2021). Análisis de series de tiempo de contaminantes atmosféricos en la ciudad de Bogotá a partir del desarrollo de modelos estadísticos ARIMA [Trabajo de grado, Universidad Distrital Francisco José de Caldas]. Repositorio UDFJC.  
<https://repository.udistrital.edu.co/server/api/core/bitstreams/93e28cb0-d9e8-4f54-a74c-08ec31078665/content>
- Secretaría Distrital de Ambiente. (s. f.). Estaciones de monitoreo de calidad del aire. Red de Monitoreo de Calidad del Aire de Bogotá.  
<http://rmcab.ambientebogota.gov.co/home/text/150>
- Yepes, C. (2020). Estimación de los niveles futuros de material particulado en el aire de Medellín y su área metropolitana [Tesis de maestría, Universidad de Antioquia].

Biblioteca Digital UdeA.

<https://bibliotecadigital.udea.edu.co/server/api/core/bitstreams/d181c2b6-89fd-4a03-9762-fbdff4667b6f/content>

Zhang, Y., & Zhao, Y. (2023). Time series prediction based on LSTM-Attention-LSTM model. *IEEE Access*, 11, 3276628. <https://doi.org/10.1109/ACCESS.2023.3276628>