

Análisis de PM2.5 suspendidos en el noroeste de Hermosillo.

DOMÍNGUEZ HURTADO JOSÉ MARTÍN¹, CIRETT GALÁN FEDERICO MIGUEL², TORRES PERALTA RAQUEL³.

RESUMEN

En la ciudad de Hermosillo, Sonora, se han empleado sensores de bajo costo para capturar datos sobre la contaminación por partículas PM2.5 y otros contaminantes atmosféricos. Dado que estos contaminantes han sido objeto de estudio durante las últimas décadas, es fundamental prever su comportamiento futuro. En este estudio, se utilizan modelos de aprendizaje automático para la predicción y el análisis de tendencias en los niveles de PM2.5. Los resultados preliminares indican que las concentraciones de contaminantes presentan una clara variabilidad estacional. La metodología propuesta sigue un enfoque sistemático para la preparación y análisis de datos en el contexto de los algoritmos de aprendizaje automático. Este enfoque incluye procesos de limpieza, exploración, tratamiento de valores atípicos y faltantes, escalado de datos categóricos, selección de características, y la partición de los datos en conjuntos de entrenamiento y prueba.

Palabras clave: Calidad del aire, PM2.5, Estacionalidad, Predicción

¹Mtro. en Internet de las Cosas e Inteligencia Artificial, Academia de Mecatrónico, Universidad Estatal de Sonora, Navojoa, Sonora, México. jose.dominguez.hurtado@ues.mx, <https://orcid.org/0009-0005-5232-4642>

²Dr. Ciencias Computacionales, Departamento Ingeniería Industrial, Universidad de Sonora, Hermosillo, Sonora, México. federico.cirett@unison.mx <https://orcid.org/0000-0003-4477-4430>

³Dra. Ciencias Computacionales, Departamento Ingeniería Industrial, Universidad de Sonora, Hermosillo, Sonora, México. raquel.torres@unison.mx <https://orcid.org/0000-0001-5836-5975>

Autor de Correspondencia: Raquel Torres Peralta, raquel.torres@unison.mx

Recibido: 14 / 11 / 2025

Aceptado: 20 / 03 / 2025

Publicado: 04 / 05 / 2025

Cómo citar este artículo:

Domínguez Hurtado, J. M., Cirett Galán, F. M., & Torres Peralta, R. (2025). Análisis de PM2.5 suspendidos en el noroeste de Hermosillo. *EPISTEMUS*, 19(38), e3809410. <https://doi.org/10.36790/epistemus.v19i38.410>

Analysis of PM2.5 Suspended in Northwest Hermosillo.

ABSTRACT

In Hermosillo, Sonora city, low-cost sensors have been used to capture data on PM2.5 particle pollution and other atmospheric pollutants. Since these pollutants have been studied over the past decades, it is essential to predict their future behavior. This study uses machine learning models for predicting and analyzing trends in PM2.5 levels. Preliminary results indicate that pollutant concentrations show clear seasonal variability. The proposed methodology follows a systematic approach for data preparation and analysis in the context of machine learning algorithms. This approach includes processes such as cleaning, exploration, handling outliers and missing values, scaling categorical data, feature selection, and partitioning the data into training and testing sets.

Key words: Air Quality, PM2.5, Stationarity, Prediction





INTRODUCCIÓN

Este estudio tiene como objetivo identificar patrones en la calidad del aire en la zona noroeste de Hermosillo, específicamente en un área donde se ha instalado un sensor, en la colonia Pueblitos. Estudios previos realizados por Meza-Figueroa [1], Santo Romo [2] y Cruz Campa, et al. [3] han señalado inconsistencias en los indicadores de calidad del aire registrados en esta área, afectados por partículas de materia, polen, NOx, SO2 y metales pesados. La zona de estudio es un área urbanizada que cuenta con locales comerciales ubicados a un costado de la acera, un campo abierto, puestos de comida en la parte inferior, y una escuela primaria situada en el lado opuesto de la calle.

Durante el día, se observa que la zona es altamente transitada, con una notable afluencia de personas caminando, automóviles y autobuses escolares. La presencia de estos autobuses genera un olor a diésel quemado, perceptible al olfato, lo que sugiere una atmósfera contaminada.

Durante el día, se observa que la zona es altamente transitada, con una notable afluencia de personas caminando, automóviles y autobuses escolares. La presencia de estos autobuses genera un olor a diésel quemado, perceptible al olfato, lo que sugiere una atmósfera contaminada. En las inmediaciones de la escuela, se encuentra un vecindario caracterizado por un constante flujo vehicular. El tráfico en la zona es causado principalmente por el transporte público, cuya alta frecuencia de paso contribuye a la liberación de contaminantes e intensifica el mal olor. Por la noche, se mantiene la actividad vehicular y peatonal, en un entorno con un estilo de vida diferente al turno matutino, donde predominan los puestos de comida, los campos deportivos y los locales comerciales.

ARIMAH ha instalado un sensor en la zona noroeste de Hermosillo. Esta área presenta niveles de partículas suspendidas PM 2.5 que superan la norma mexicana, según estudios anteriores que han evidenciado el deterioro en la calidad del aire. En este documento, se realiza un análisis de un conjunto de datos recolectados por un sensor de PM2.5, como temperatura, humedad, presión, entre otros. Los datos fueron recolectados durante un año, con una frecuencia de 5 minutos, y se



utilizaron para predecir y detectar patrones en la distribución de contaminantes. Para ello, se aplicaron modelos de predicción como ARIMA, SARIMA y Prophet.

PLANTEAMIENTO DEL PROBLEMA

En el contexto de la ciudad de Hermosillo, se busca identificar y comprender los contaminantes presentes en el aire, a través de su ubicación geográfica, hora y fecha de detección. Con este propósito, se ha establecido una red de sensores distribuidos en diversas zonas de la ciudad. El objetivo principal de esta red es analizar los patrones de los principales agentes contaminantes, que contribuyen a la generación de residuos en el aire.

OBJETIVO

El objetivo de este estudio es detectar y predecir, mediante un algoritmo de inteligencia artificial, ciclos en los comportamientos de los principales contaminantes del aire en la zona noroeste en la ciudad de Hermosillo, Sonora, de acuerdo con las lecturas de un sensor ubicado en la zona de Pueblitos que mide la calidad del aire.

ESTADO DEL ARTE.

Zhou et al. [4] realizaron un análisis sistémico de 118 estudios que aplican Deep Learning (DL) para la predicción de concentraciones de PM2.5. Utilizando los lineamientos PRISMA, los autores categorizaron los modelos existentes, destacaron siete tipos de estructuras de DL y proporcionaron una visión general del estado actual de la investigación. Además, introdujeron un marco de evaluación estandarizado para mejorar la comparabilidad y fiabilidad de los resultados futuros, lo que representa un avance importante hacia la estandarización en este campo.

En Lahore, Pakistán, Bhatti et al. [5] analizaron la evolución de los principales contaminantes atmosféricos, incluyendo PM10, PM2.5, SO₂, CO y NO₂, entre los años 2014 y 2019. Observaron una fuerte correlación entre estos contaminantes, junto con variaciones estacionales





significativas, lo que sugiere la necesidad de medidas de control continuas para mejorar la calidad del aire en la región.

Por otro lado, Mohammadi et al. [6] implementaron un modelo de redes neuronales artificiales (ANN) para la predicción de concentraciones de PM_{2.5}. Utilizando la herramienta Neural Net Toolbox de Matlab, aplicaron una red de retropropagación multicapa, entrenada con el 70% de los datos, validada con un 15% y probada con el 15% restante. Este enfoque permitió obtener predicciones precisas utilizando el algoritmo de retropropagación Levenberg–Marquardt.

En la ciudad de Beijing, Pak et al. [7] propusieron el uso de redes neuronales convolucionales (CNN) y memoria a corto y largo plazo (LSTM) para la predicción de PM_{2.5}. Estos modelos proporcionaron predicciones de alta precisión, por lo que son herramientas fundamentales para la gestión ambiental y la toma de decisiones gubernamentales en relación con la calidad del aire.

Con un enfoque novedoso, Xiao et al. [8] desarrollaron el modelo WLSTME (memoria de red neuronal de pesos a largo y corto plazo), diseñado para capturar las correlaciones espaciotemporales de los contaminantes del aire. Este modelo consideró factores como la densidad del aire y las condiciones del viento, y demostró un rendimiento superior con los menores errores RMSE y MSE, lo que lo posiciona como un método prometedor en la predicción de contaminación atmosférica.

Adicionalmente, Jin et al. [9] analizaron la influencia de factores ambientales en la distribución de PM_{2.5}, tales como el índice de vegetación normalizado, la precipitación, la temperatura, la velocidad del viento y la elevación. Su estudio reveló cómo estos factores impactan las concentraciones de PM_{2.5} en diferentes regiones de China. Asimismo, subrayó la complejidad de las interacciones entre variables meteorológicas y contaminantes.

Para abordar los vacíos de datos, Xiao et al. [10] emplearon estrategias de relleno de huecos y aplicaron modelos de predicción tipo Random Forest. Integraron observaciones terrestres,



simulaciones de modelos de transporte químico (CTM) y datos satelitales (AOD). Generaron predicciones diarias de PM2.5 con una resolución espacial de 1 km para áreas clave como Beijing-Tianjin-Hebei.

Ban y Shen [11] presentaron un enfoque híbrido que combina el algoritmo CEEMDAN (descomposición empírica en ensamble con ruido adaptativo) con modelos de DL. Este método descompone los datos de contaminantes en diferentes componentes antes de ser procesados por el modelo, lo que mejora significativamente la precisión de las predicciones.

En Taiwán, Chen y Li [12] destacaron la dificultad de aplicar métodos convencionales a datos meteorológicos de alta dimensionalidad, lo que afecta la precisión de las predicciones de PM2.5. Subrayan la necesidad de técnicas avanzadas que puedan identificar patrones complejos en los datos para mejorar la precisión y, por ende, la salud pública.

Ma et al. [13] lograron una mejora del 50% en la precisión de las predicciones de calidad del aire en Shanghai mediante la integración de datos meteorológicos en el modelo XGBoost, superando a métodos anteriores como WRF-Chem. Este enfoque demuestra la capacidad de los modelos Machine Learning (ML) para ofrecer predicciones más exactas en áreas urbanas.

METODOLOGIA

Los datos recabados en este estudio se proporcionaron a la Red de Monitoreo Ambiental de Hermosillo (REMAH), que es producto del trabajo en conjunto de académicos y estudiantes de la Universidad de Sonora y la Agencia Municipal de Energía y Cambio Climático (AMECC). El proyecto es parte de las actividades del grupo TEAMS (Grupo de Transporte Eléctrico Avanzado y Movilidad Sostenible) que tiene como propósito la mejora en la calidad del aire en las ciudades de Hermosillo y Cuernavaca.





Actualmente, la REMAH cuenta con 12 sensores UBISmart+ colocados en forma distribuida en toda la ciudad de Hermosillo. Para este estudio, solo se utilizará información generada por el sensor que se encuentra localizado en la colonia Pueblitos, al noroeste de la ciudad cabe aclarar que UbiSmart AQM+ es un pequeño sensor que mide, monitorea, registra, analiza y comunica datos precisos sobre la calidad del aire local y el medio ambiente. Montado en un poste de luz a aproximadamente 4 metros del suelo, los sensores muestrean el aire circundante y reportan la información a la plataforma basada en la nube.

Los datos son recolectados cada 5 minutos y se guardan localmente cada 10 minutos en la aplicación UbiVu. Posteriormente, los datos son enviados a la red de UbiCu por un servidor dentro de la Universidad de Sonora. El cómo funcionan los sensores; después de la instalación, el sistema mide la calidad del aire correspondiente, partículas PM, temperatura, presión, humedad y datos de ruido. Estos datos se envían a UbiVu, a través de la red 4G.

Para el procesamiento de datos, limpiar datos para un contexto de algoritmos de aprendizaje automático implica usar bibliotecas necesarias para cambiar los datos de un marco de datos; explorarlos, tratarlos con valores faltantes y valores atípicos, codificando diferentes valores, escalabilidad categórica; elegir los más relevantes y, finalmente, dividir los datos en conjuntos de prueba y entrenamiento.

Al obtener los datos procesados hay que utilizar un espacio para su visualización, se requiere una inspección profunda de los datos, que permita detectar patrones, tendencias y relaciones que pueden ser cruciales para tomar decisiones. Para ello, es necesario utilizar herramientas para crear gráficos, estadísticas descriptivas y visualización que facilitan la comprensión de los datos.

La segmentación de datos es un proceso analítico para mantener un conjunto de grupos y subgrupos más pequeños, según algunas características y puntos compartidos. Se realizaron varias pruebas para comprobar cuál es el comportamiento de los datos, en este estudio se realizaron pruebas de normalidad de Shapiro-Wilk, y Kolmogórov-Smirnov y para la prueba de estacionalidad de Augmented Dickey Fuller. Utilizando el algoritmo de aprendizaje automático, se utilizaron datos de entrenamiento para aprender patrones y correlaciones. Esto es esencial



durante estas fases, para minimizar la diferencia entre la predicción del modelo y los valores reales. Los datos se analizaron por medio de diferentes modelos, ARIMA, SARIMA y Prophet.

RESULTADOS

Para comprender la historia que nos cuentan los datos, realizamos una gráfica de series de tiempo con los datos obtenidos del sensor de PM2.5 sin preprocesar. Se utilizaron datos totales obtenidos durante un año en el área de Pueblitos, relacionado con la fecha y la cantidad de partículas PM2.5 encontradas en el área. Se identificaron múltiples valores atípicos en los datos. Para mitigar su impacto en el análisis, se procesó la serie temporal eliminando aquellos valores que excedían tres desviaciones estándar.

En la Figura 1, se presenta el gráfico con los datos ya procesados, donde se eliminan aquellos que podrían distorsionar los resultados. Al comparar estos datos con las fechas de recolección, se observa una variabilidad en el comportamiento, donde se destaca que el periodo comprendido entre septiembre de 2022 y mayo de 2023 muestra las concentraciones más elevadas en comparación con el periodo de julio a noviembre de 2023.



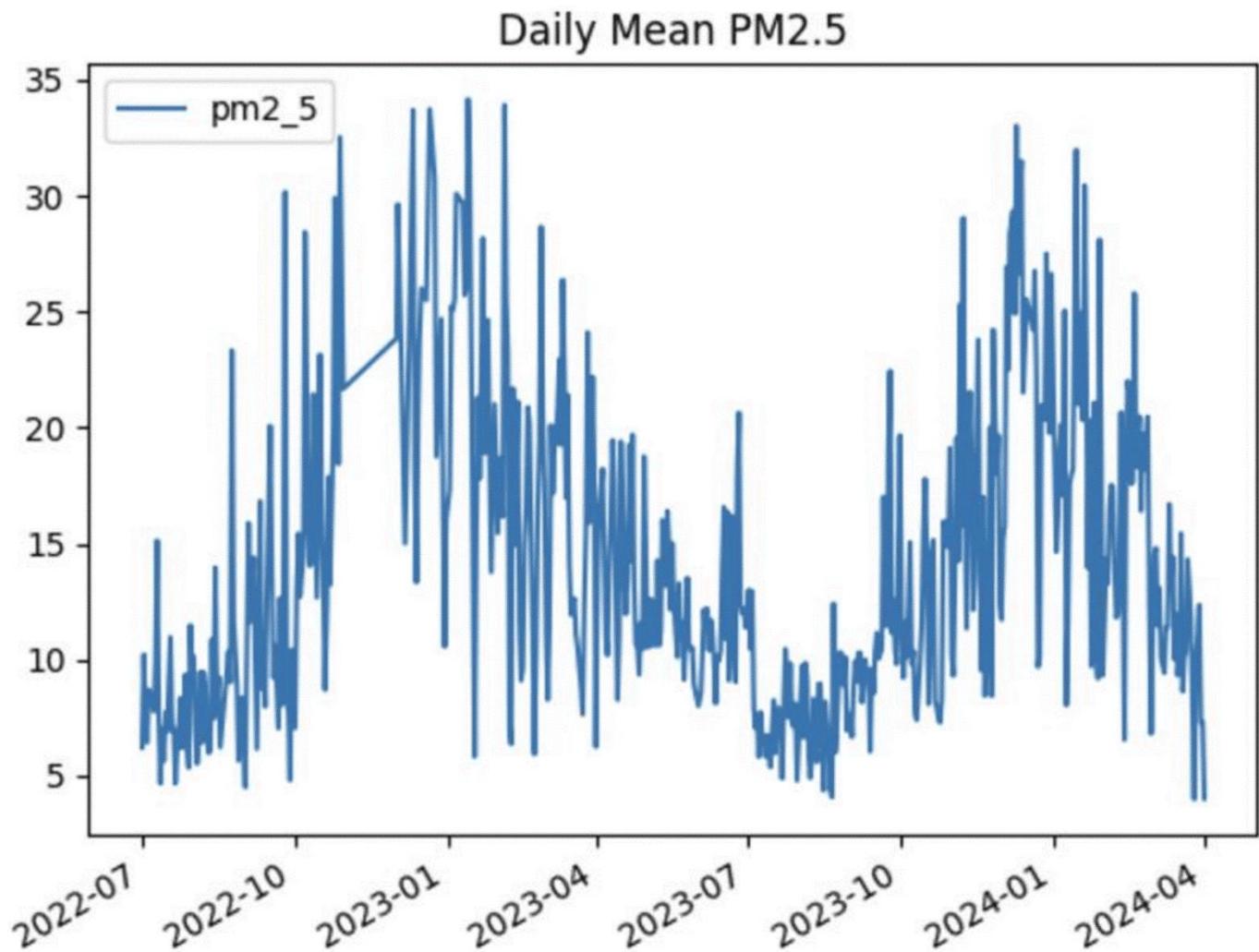


Figura 1. Gráfico de series de tiempo de PM 2.5 sin valores atípicos del área de Pueblitos en la ciudad de Hermosillo, Sonora.

El comportamiento a lo largo de los días es muy irregular, con picos altos cada 3 a 5 días, con un comportamiento sinusoidal que aumenta y disminuye. Si se observa de cerca el comportamiento de los datos en la Figura 2, este se puede apreciar durante los días de enero de 2023.

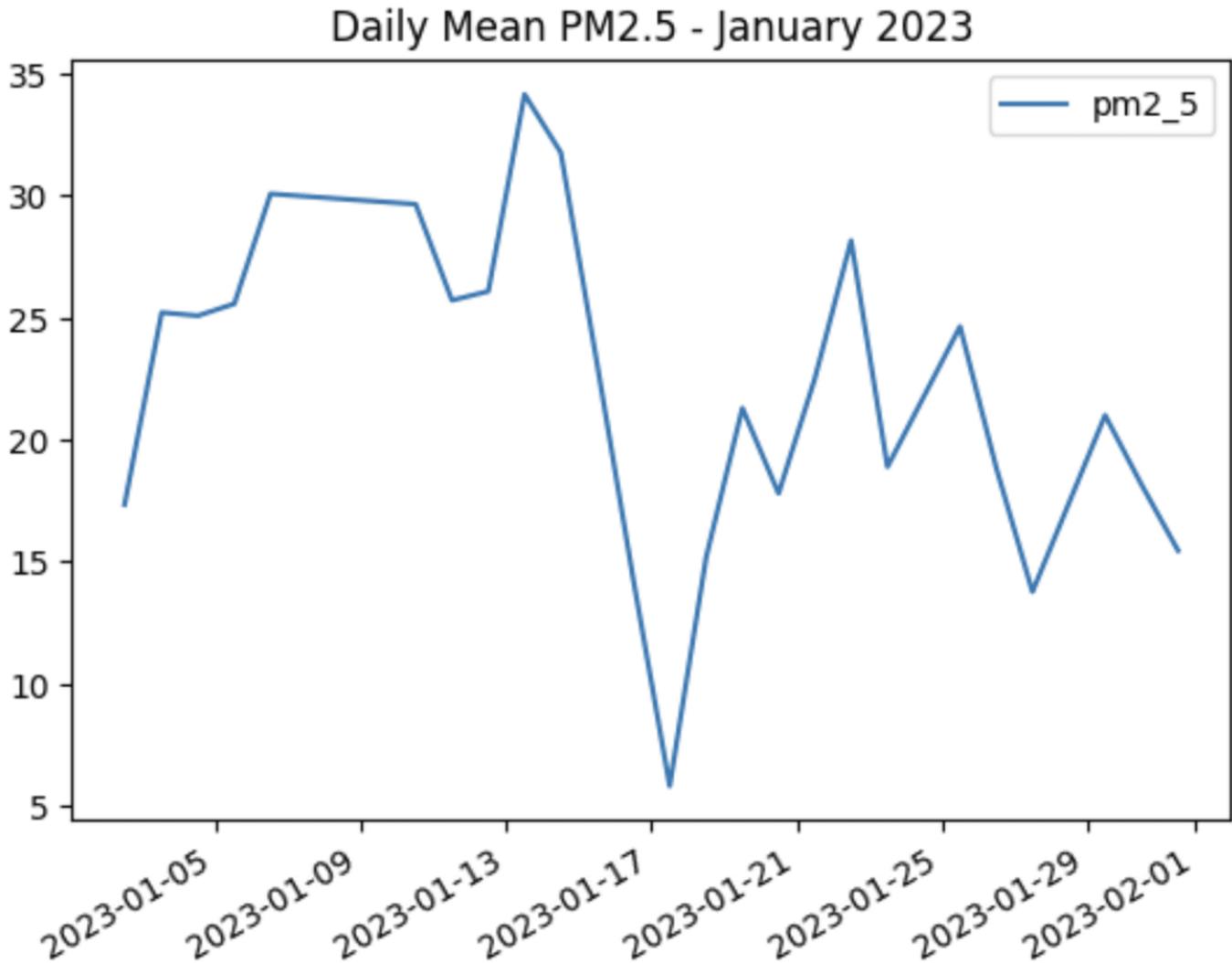


Figura 2. Gráfico de datos limpios para el mes de enero en el área de Pueblitos en la ciudad de Hermosillo, Sonora.

Un gráfico Q-Q (cuantil-cuantil) es una herramienta gráfica utilizada para comparar la distribución de dos conjuntos de datos. Se crea trazando los cuantiles de un conjunto frente a los cuantiles del otro. Si ambos conjuntos de datos tienen la misma distribución, los puntos en el gráfico Q-Q caerán en una línea recta.



En nuestro análisis, empleamos un gráfico Q-Q para evaluar la distribución de los datos obtenidos. En la Figura 3, observamos que los puntos se desvían de la línea recta en varias posiciones, lo que indica que los datos no siguen una distribución estándar. Esto sugiere la necesidad de calibrar los parámetros del modelo. Determinar adecuadamente la calidad y cantidad de datos para el entrenamiento es crucial para obtener mejores resultados y aumentar la efectividad del proceso.

Específicamente, podemos observar que, los puntos correspondientes a los valores más pequeños están por debajo de la línea recta, lo que indica que hay menos valores pequeños en los datos de lo que se esperaría si estuvieran distribuidos normalmente. Los puntos correspondientes a los valores más grandes están por encima de la línea recta, lo que indica que hay más valores grandes en los datos de lo que se esperaría si estuvieran distribuidos normalmente.

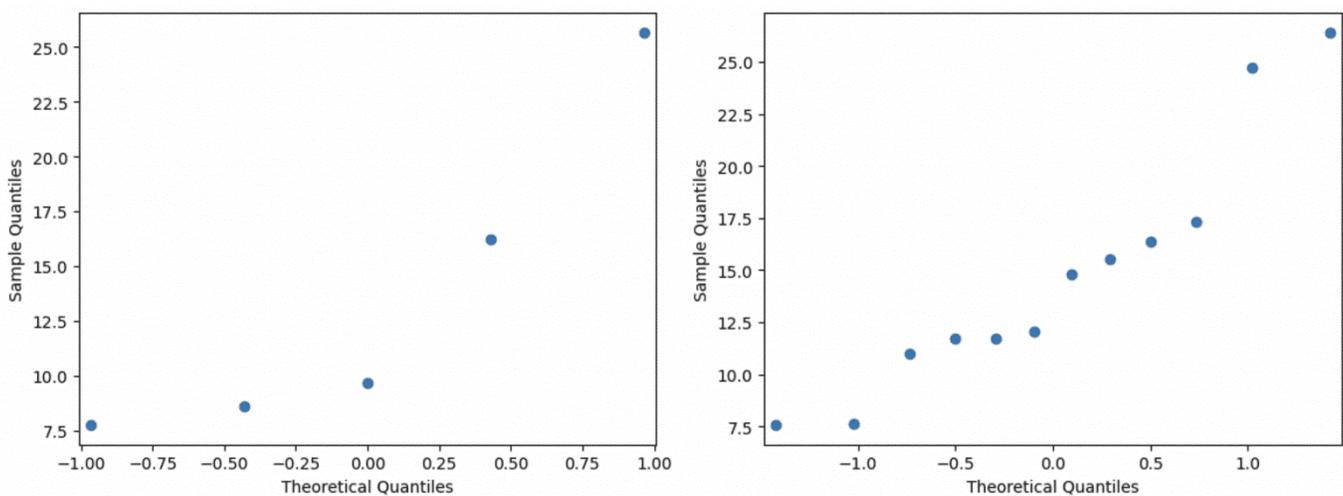


Figura 3. Gráfico de Cuantil-Cuantil de datos en los años 2022 y 2023 de Pueblitos en la ciudad de Hermosillo, Sonora.

El gráfico Q-Q indica que los residuos no siguen una distribución estándar. Esto puede deberse a la presencia de valores atípicos, heterocedasticidad o no linealidad en los datos. Estos hallazgos sugieren que los residuos no son normales y que el modelo de regresión lineal podría no ser adecuado para este conjunto de datos. Para realizar predicciones, se debe utilizar un conjunto de

datos con pocas irregularidades, lo que facilita visualizar el comportamiento de la variable PM2.5. En este caso, los datos con menos ruido han sido los promedios mensuales y semanales del año, lo que permite comparaciones y predicciones más efectivas. Los datos promedio semanales para el año 2023 pueden contener irregularidades en comparación con los datos semanales del 2022, que forman casi una línea recta. Esto indica falta de datos normativos.

Para identificar puntos de datos irregulares, se utilizó un gráfico de boxplot, que ayuda a mostrar la distribución de los datos, en los que se observa una pendiente que está casi centrada en el eje x, con dos puntos de datos atípicos. Esto se interpreta en la Figura 4.b. Lo anterior contrasta con los datos de 2022, que muestran un comportamiento dentro de los rangos y con una ligera inclinación hacia la izquierda en el eje x, como se demuestra en la línea centrada en el boxplot de la Figura 4.a.

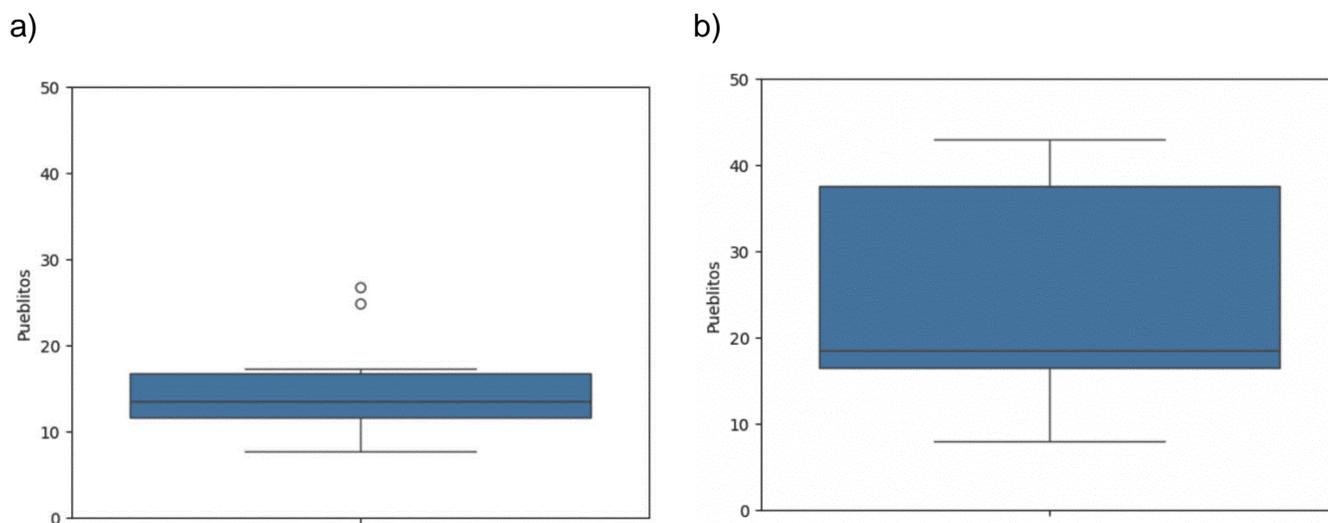


Figura 4. Gráfico de caja de datos mensuales promedio en el área de Pueblitos de Hermosillo, Sonora. En la figura a) se muestran los registros de 2022, en la figura b) se muestran los registros de 2023.



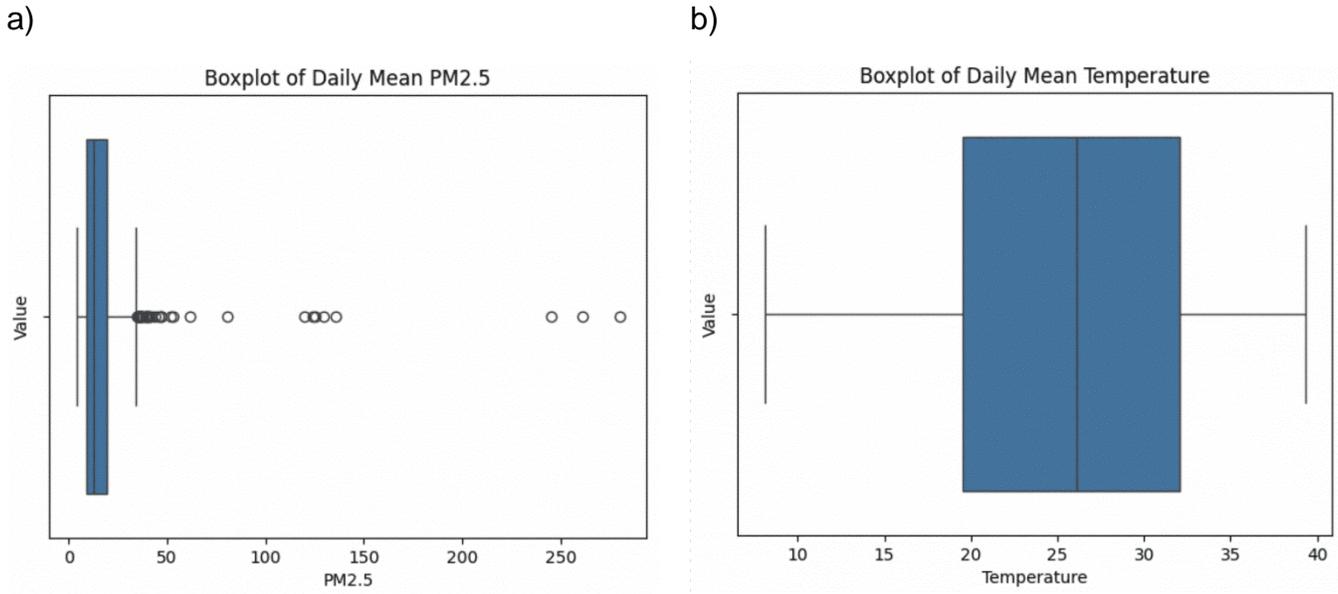


Figura 5. Gráfico de comparación de caja entre PM2.5 y temperatura.

Dado que el gráfico de boxplot en la Figura 5 no muestra datos fuera del rango, no es necesario buscar valores atípicos. Para hacerlo con los datos de 2023, es necesario calcular la media y la desviación estándar. Utilizando algunas funciones de la biblioteca numpy en Python, fue posible recuperar los resultados en la Tabla 1.

Tabla 1. Promedio y desviación estándar del conjunto de datos obtenidos de 2023.

Mean	14.83
Standard Deviation	5.73

La Figura 6 muestra un gráfico con el conjunto de datos en el que se han eliminado los valores atípicos, lo que resulta en un gráfico que visualiza mejor la interpretación realizada con el boxplot.



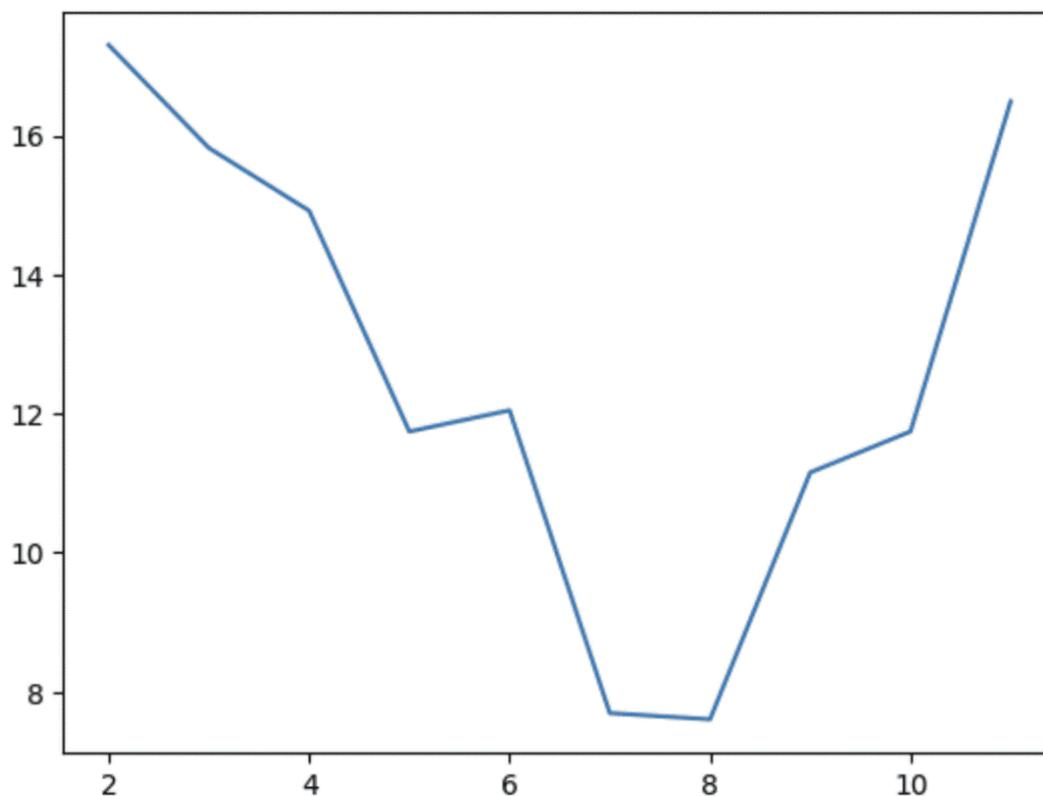


Figura 6. Registros de PM2.5 mensuales promedios de PM2.5 para el año 2023 en el área de Pueblitos de Hermosillo, donde se eliminaron puntos de datos atípicos. El eje x son los meses del año y el eje y, las mediciones de partículas suspendidas.

Realizando pruebas de Kolmogorov-Smirnov y Shapiro-Wilkos, datos promediados de 2023 muestran una distribución normal. Así, se confirma la hipótesis nula, indicando que no es necesario realizar ningún cambio. Con la ayuda de gráficos separados para cada variable, se puede realizar un análisis del comportamiento de cada valor, considerando que el sensor está ubicado en una zona urbana con un clima seco en la ciudad de Hermosillo.

Observamos que los valores de presión y humedad aumentan durante la temporada de invierno de diciembre a marzo. Cuando llegan la primavera y el verano, estos valores disminuyen, posiblemente debido a factores climáticos o comportamiento social durante estas estaciones.



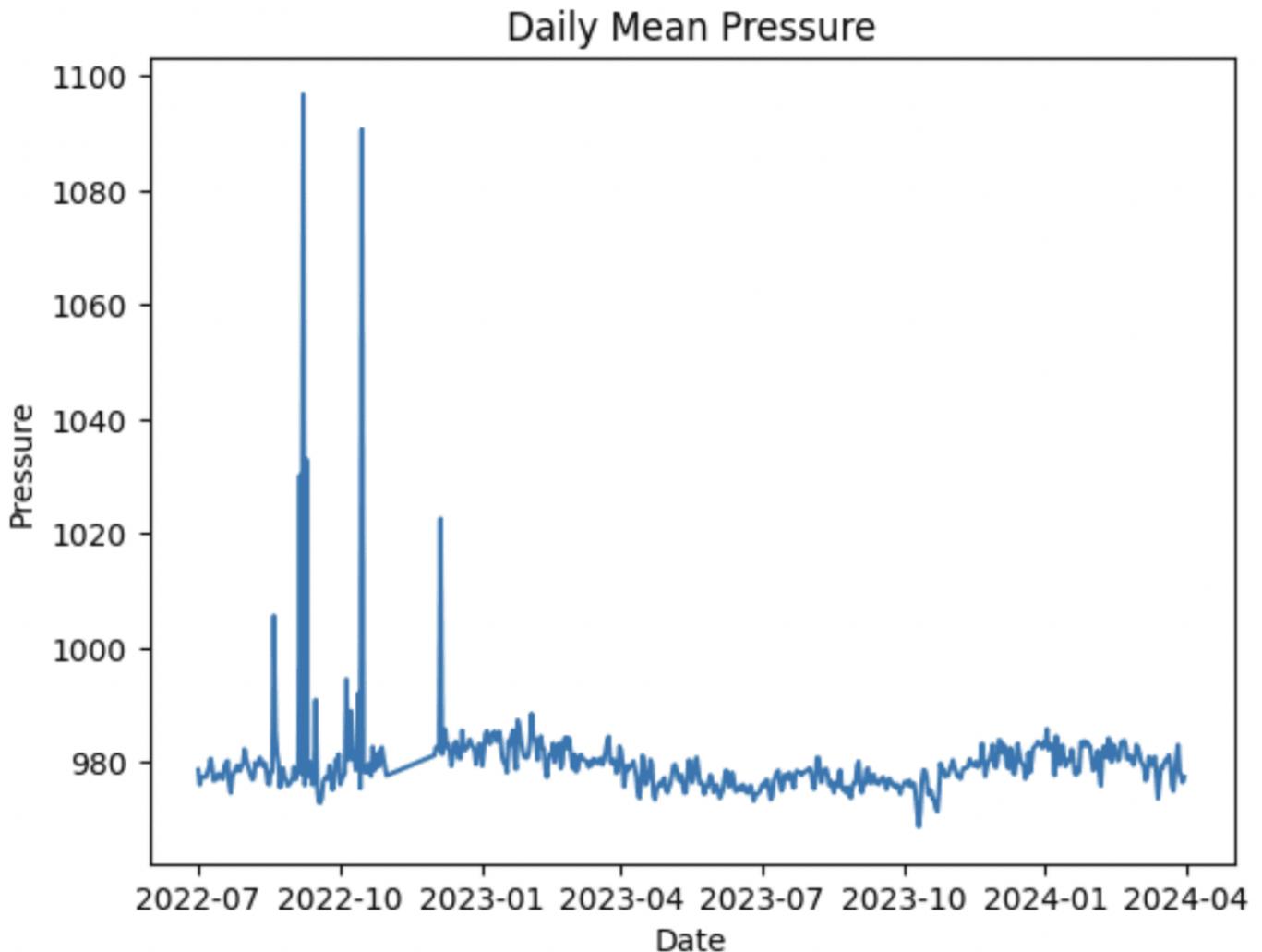


Figura 7. Gráficos de datos históricos de presión, promediados por día.

Se muestra una gráfica de series de tiempo de los niveles de PM2.5 registrados por el sensor ubicado en la zona noroeste de Hermosillo, utilizando los datos sin preprocesar. Esta gráfica permite visualizar cómo varían las concentraciones de partículas finas PM2.5 a lo largo del tiempo, revelando patrones, tendencias y fluctuaciones en la calidad del aire de la zona estudiada.

Es notable que el comportamiento demuestra una relación que puede ser menos visible para algunas variables, pero más evidente para otras, aumentando y disminuyendo como se mencionó en el análisis anterior.

Para realizar una prueba de normalidad y estacionalidad, es necesario utilizar un algoritmo para eliminar datos atípicos, logrando así mejores resultados. Con el uso de las pruebas Shapiro-Wilk y Kolmogórov-Smirnov se puede confirmar que no se tiende a una distribución normal, lo cual requiere implementar técnicas estadísticas paramétricas o retomar transformaciones adicionales para un mejor análisis. Esta conducta puede estar influida por comportamientos durante la obtención de los datos.

La estacionalidad de los datos demuestra que se tiende a un patrón estacional, ya que estos varían regularmente durante el año. Esta conducta ayuda a entender la dinámica temporal del comportamiento del contaminante recolectado y debe considerarse para reducir la vulnerabilidad. La prueba de estacionalidad muestra un resultado similar para todas las variables, rechazando la hipótesis nula e indicando que los datos son estacionarios, como se muestra en la Tabla 2.

Tabla 2. Pruebas de normalidad y estacionalidad de valores de PM2.5

Variable	Prueba de normalidad				Prueba de estacionalidad	
	Kolmogorov-Smirnov	P-Value	Shapiro-Wilk	P-Value	Augmented Dickey Fuller	P-Value
PM2.5	0.14	0.00	0.88	0.00	-10.73	0.00
Presión	0.05	0.00	0.99	0.00	-7.06	0.00
Humedad	0.07	0.00	0.96	0.00	-8.75	0.00





	Temperatura	0.03	0.00	0.99	0.00	-3.61	-0.01
--	-------------	------	------	------	------	-------	-------

Para verificar qué modelo puede dar mejores resultados al predecir el comportamiento de los datos, se ha realizado el procedimiento de Mean Square Error. Los resultados se muestran en la Tabla 3. Se puede afirmar que el modelo con mejor resultado es el de Prophet.

Tabla 3. Tasa de error para cada técnica de predicción utilizada.

Modelo	Error cuadrático medio
Prophet	20.27
SARIMA	20.68
ARIMA	29.88

Las Figuras 8 y 9 muestran los resultados de aplicar los modelos ARIMA y SARIMA, respectivamente, utilizando los datos previamente analizados. Se aplica un algoritmo de aprendizaje automático para prever el comportamiento futuro de los datos, y estas predicciones se comparan con los datos reales.



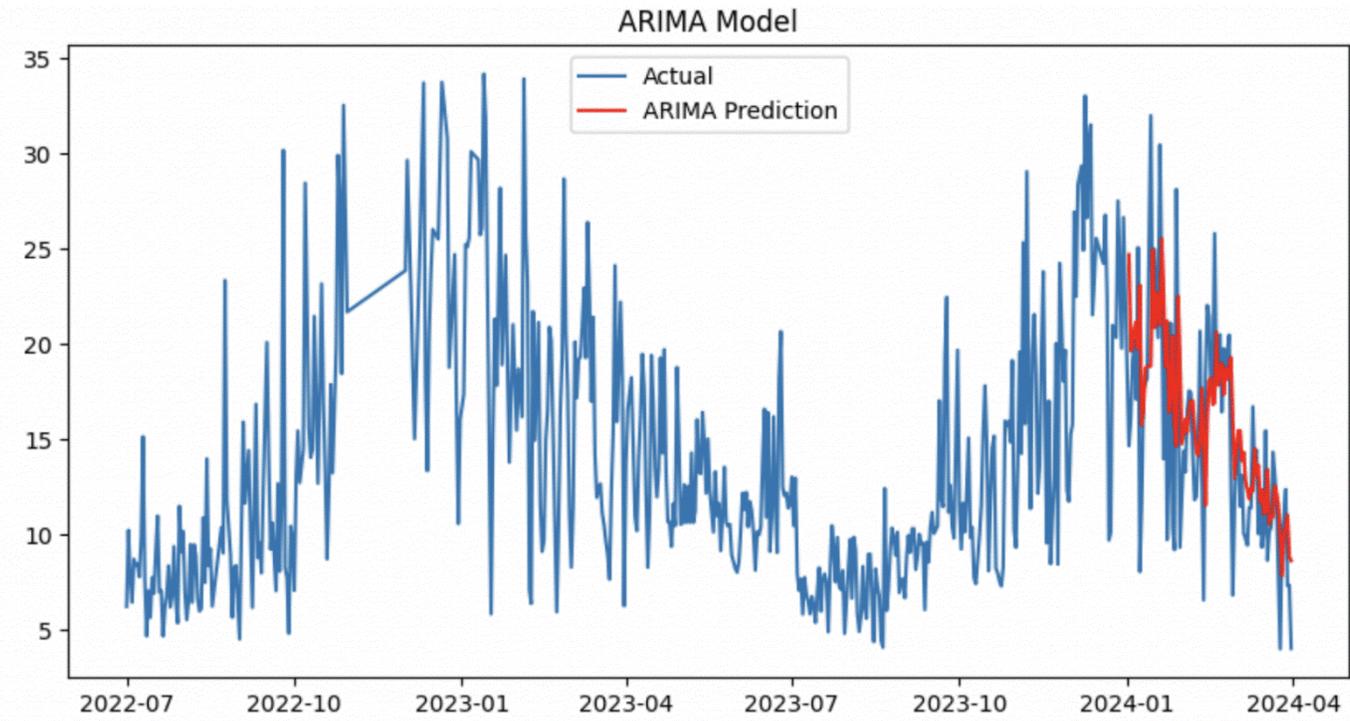


Figura 8. Gráfico del modelo ARIMA con datos actuales y predichos de PM2.5.

Los modelos ARIMA y SARIMA difieren al incorporar un componente estacional, lo que proporciona una diferencia significativa. Este componente estacional utiliza patrones estacionales comunes para predecir datos regulares, como la temperatura mensual, por ejemplo.



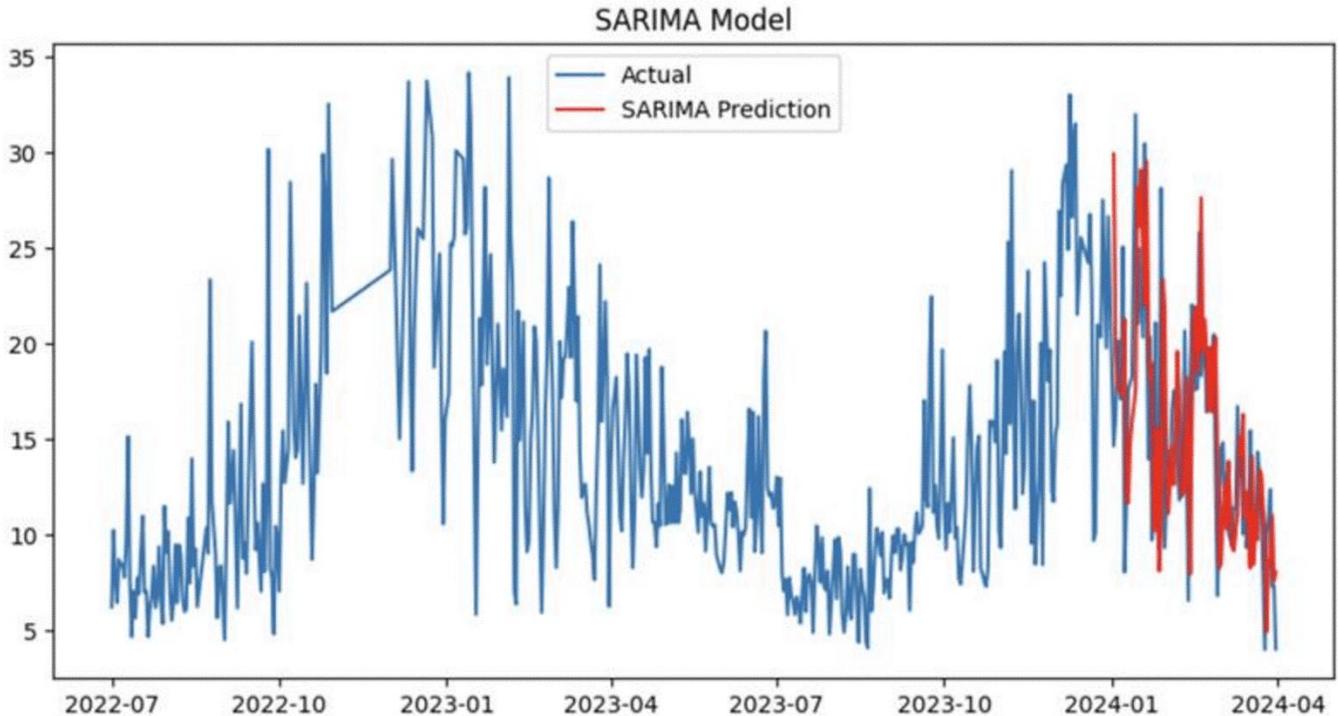


Figura 9. Gráfico del modelo SARIMA con datos actuales y predichos de PM2.5.

Comparando los datos de estacionalidad con el modelo, utilizando series temporales, los resultados son muy precisos, lo que resulta en un gráfico con líneas que coinciden con los datos actuales. La Figura 10 muestra el resultado de Prophet, que fue utilizado para obtener un menor error debido a que ARIMA no tiene en cuenta estacionalidad; mientras que SARIMA y Prophet sí lo hacen. SARIMA lo hace mediante términos estacionales explícitos, y Prophet mediante componentes aditivos. Prophet es más flexible en términos de manejo de datos no estacionarios, faltantes y cambios abruptos en la tendencia, mientras que ARIMA y SARIMA requieren datos estacionarios y pueden ser más sensibles a estos problemas.

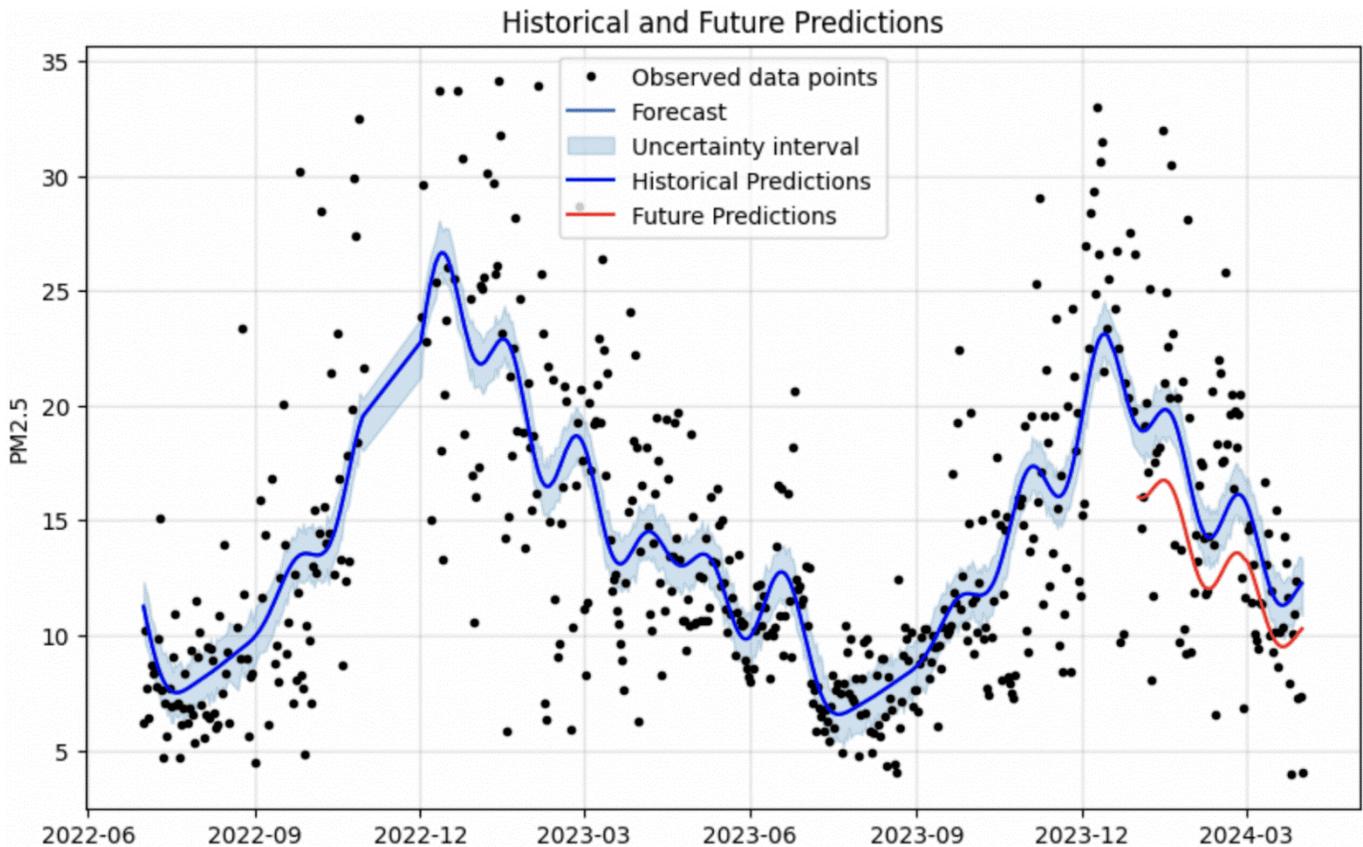


Figura 10. Gráfico del modelo de Prophet con datos históricos y predichos de PM2.5

DISCUSIÓN

El análisis y predicción de la contaminación del aire es un tema de gran relevancia en el ámbito de la salud pública y la gestión ambiental, especialmente en regiones donde la exposición a material particulado fino (PM2.5) representa un riesgo significativo. En este contexto, el presente estudio se enfoca en la detección y predicción de la calidad del aire en la ciudad de Hermosillo, Sonora, mediante la implementación de sensores de bajo costo y modelos de aprendizaje automático. La combinación de estas tecnologías permite una caracterización más accesible y detallada de los niveles de contaminación atmosférica, lo que facilita la identificación de patrones temporales y la implementación de estrategias de mitigación adecuadas.





Para garantizar la fiabilidad de los resultados obtenidos, se llevó a cabo un proceso metodológico riguroso que incluyó la limpieza, exploración y visualización de los datos recolectados. Estas etapas fueron fundamentales para detectar patrones significativos y posibles anomalías en las mediciones. Posteriormente, se aplicaron diversos algoritmos predictivos especializados en el análisis de series temporales con el fin de modelar la evolución de la contaminación del aire. Entre los modelos empleados se encuentran ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal ARIMA) y Prophet, los cuales han demostrado ser herramientas efectivas en la predicción de fenómenos ambientales con comportamiento estacional y tendencias a largo plazo.

El desempeño de cada modelo fue evaluado utilizando métricas estadísticas de error con el objetivo de determinar cuál de ellos presentaba mayor precisión en la estimación de los niveles futuros de PM2.5. Como criterio principal de comparación se empleó el error cuadrático medio (RMSE, por sus siglas en inglés), el cual permite cuantificar la desviación promedio entre los valores predichos y las mediciones reales. Los resultados obtenidos evidenciaron que el modelo Prophet presentó el menor margen de error entre los algoritmos evaluados, lo que sugiere que es el más adecuado para la predicción de la calidad del aire en la región de estudio.

Estos hallazgos destacan la importancia de la aplicación de modelos avanzados de aprendizaje automático para mejorar la vigilancia y gestión de la contaminación atmosférica. La implementación de sensores de bajo costo, en conjunto con herramientas analíticas robustas, representa una estrategia viable para optimizar la toma de decisiones en políticas ambientales y reducir la exposición de la población a contaminantes nocivos. Futuros estudios podrán enfocarse en la integración de variables meteorológicas adicionales y en el desarrollo de modelos híbridos que permitan mejorar aún más la precisión de las predicciones y la identificación de factores determinantes en la variabilidad de la calidad del aire.

CONCLUSIONES

Se emplearon diversos modelos de predicción de series de tiempo con el propósito de analizar y proyectar las temporadas de mayor concentración de material particulado fino (PM2.5) en la



atmósfera. Los resultados obtenidos indican que las concentraciones más elevadas de este contaminante ocurren predominantemente durante las estaciones de otoño e invierno. Además, se estima que, en ausencia de intervenciones efectivas, estos episodios de alta contaminación continuarán manifestándose en el futuro con una distribución estacional similar.

En este sentido, se resalta la importancia de la implementación de políticas ambientales más rigurosas y sostenibles que contribuyan a la mejora de la calidad del aire y a la protección de la salud pública, especialmente en aquellas estaciones del año en las que se registran las mayores concentraciones de PM2.5. Asimismo, el análisis de los datos ha permitido identificar un patrón estacional en la variabilidad de las concentraciones de PM2.5. Este patrón presenta un valor normativo específico que refleja fluctuaciones regulares a lo largo del año. La identificación y comprensión de esta dinámica temporal resulta fundamental para el diseño e implementación de estrategias de mitigación eficaces, ya que permite una mejor planificación de medidas preventivas y correctivas orientadas a la reducción de la exposición a este contaminante atmosférico.

REFERENCIAS

- [1] D. Meza-Figueroa. «Heavy metal distribution in dust from elementary schools in Hermosillo, Sonora, México». *Atmospheric Environment*, 2007, 276-28.
- [2] A. Santos Romo. *Evaluation of air quality for PM10, NO_x and SO₂ suspended particles in the city of Hermosillo, Sonora*, 2006.
- [3] M. E. Cruz Campas, A. Gómez Alvarez, Quintero Núñez, M., y Varela Salazar, J. (2013). «Evaluación de la calidad del aire respecto de partículas suspendidas totales (PST) y metales pesados (Pb, Cd, Ni, Cu, Cr) en la ciudad de Hermosillo, Sonora, México, durante un periodo anual». *Revista internacional de contaminación ambiental*, 29(4), 269-283.
- [4] S. Zhou, W. Wang, L. Zhu, Q. Qiao, y Y. Kang, «Deep-learning architecture for PM2. 5 concentration prediction: A review», *Environmental Science and Ecotechnology*, 100400, 2024. <https://doi.org/10.1016/j.es.2024.100400>
- [5] U. A. Bhatti, Y. Yan, M. Zhou, S. Ali, A. Hussain, H. Qingsong, y L. Yuan, «Time series analysis and forecasting of air pollution particulate matter (PM 2.5): an SARIMA and factor analysis approach», *Access*, 9, 41019-41031, 2021. <https://doi.org/10.1109/ACCESS.2021.3060744>





- [6] F. Mohammadi, H. Teiri, Y. Hajizadeh, A. Abdolahnejad, y A. Ebrahimi, «Prediction of atmospheric PM2.5 level by machine learning techniques in Isfahan, Iran», *Scientific Reports*, 14(1), 2109, 2024. <https://doi.org/10.1038/s41598-024-52617-z>
- [7] U. Pak, J. Ma, U. Ryu K. Ryom, U. Juh yok, K. Pak, y C. Pak, «Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China», *Science of the Total Environment*, 699, 133561, 2020. <https://doi.org/10.1016/j.scitotenv.2019.07.367>
- [8] F. Xiao, M. Yang, H. Fan, G. Fan, y M. A. Al-Qaness, «An improved deep learning model for predicting daily PM2.5 concentration», *Scientific reports*, 10(1), 20988, 2020. <https://doi.org/10.1038/s41598-020-77757-w>
- [9] H. Jin, X. Chen, R. Zhong, y M. Liu, «Influence and prediction of PM2.5 through multiple environmental variables in China», *Science of The Total Environment*, 849, 157910, 2022. <https://doi.org/10.1016/j.scitotenv.2022.157910>
- [10] Q. Xiao, G. Geng, J. Cheng, F. Liang, R. Li, X. Meng, y K. He, «Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models», *Atmospheric Environment*, 244, 117921. 2021. <https://doi.org/10.1016/j.atmosenv.2020.117921>
- [11] W. Ban y L. Shen, «PM2.5 prediction based on the CEEMDAN algorithm and a machine learning hybrid model», *Sustainability*, 14(23), 16128, 2022. <https://doi.org/10.3390/su142316128>
- [12] Y. C. Chen y D. C. Li, «Selection of key features for PM2.5 prediction using a wavelet model and RBF-LSTM», *Applied Intelligence*, 51, 2534-2555, 2021. <https://doi.org/10.1007/s10489-020-02031-5>
- [13] J. Ma, Z. Yu, Y. Qu, J. Xu, y Y. Cao, «Application of the XGBoost machine learning method in PM2.5 prediction: a case study of Shanghai», *Aerosol and Air Quality Research*, 20(1), 2020, pp. 128-138. <https://doi.org/10.4209/aaqr.2019.08.0408>

Cómo citar este artículo:

Domínguez Hurtado, J. M., Cirett Galán, F. M., & Torres Peralta, R. (2025). Análisis de PM2.5 suspendidos en el noroeste de Hermosillo. *EPISTEMUS*, 19(38), e3809410. <https://doi.org/10.36790/epistemus.v19i38.410>

