

Data science-based tool to reduce measurement errors in atmospheric monitoring systems

Jonathan González^{1, }, Eladio Quintero^{1, }, Felisindo García^{1, }, Antony García^{1, }, Yessica Sáez, Ph.D.^{1,2, }, and Edwin Collado, Ph.D.^{1,2, }*

¹ Universidad Tecnológica de Panamá, Panamá, {jonathan.gonzalez14, eladio.quintero1, felisindo.garcia, antony.garcia, yessica.saez, edwin.collado}@utp.ac.pa

² Centro de Estudios Multidisciplinarios en Ciencias, Ingeniería y Tecnología AIP (CEMCIT AIP, Panamá

*Corresponding author: edwin.collado@utp.ac.pa

Abstract – The growing concern about air pollution and climate change, derived from global industrial growth, has generated environmental disturbances with unpredictable consequences for human health. Among the most harmful polluting agents are airborne particles, known as particulate matter, whose rapid dispersion in extensive areas exacerbates the problem. To address this issue, an urban atmospheric monitoring system based on the Internet of Things was developed to measure PM_{2.5} and PM₁₀ particulate concentrations. This system primarily consists of a Nova PM SDS011 sensor to take measurements of PM₁₀ and PM_{2.5}, a DHT22 sensor to record temperature and humidity levels, and an ESP32 microcontroller programmed in MicroPython, which was responsible for data collection and transmission via WiFi. The system was validated using the AEROQUAL S500 as reference equipment, concluding that the proposed measuring station adequately measures the variables under study. However, during the validation period it was observed that part of the data collected presented errors, either because the measurement was corrupted, the system did not work correctly, and/or an unwanted value was recorded. Therefore, this work proposes to develop a Data Science-based tool to properly process the data and reduce the number of errors in the collected measurements. Specifically, this tool uses Python to apply outlier cleaning techniques based on Interquartile Range. This allows for the identification and correction of erroneous readings caused by external factors. The aim is to generate precise information on air pollution levels to support decision making. The developed prototype demonstrates that, by using the Internet of Things and simple Data Science techniques, it is possible to provide an urban atmospheric monitoring system with high precision, low cost, and reliable data, allowing a better understanding of health risks to support the adoption of environmental protection measures.

Keywords-- Air pollution, data science, particulate matter, statistical methods, sensors.

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

Herramienta basada en ciencia de datos para reducir errores en las mediciones en sistemas de monitoreo atmosférico

Jonathan González^{1, }, Eladio Quintero^{1, }, Felisindo García^{1, }, Antony García^{1, }, Yessica Sáez, Ph.D.^{1,2, }, and Edwin Collado, Ph.D.^{1,2, },*

¹ Universidad Tecnológica de Panamá, Panamá, {jonathan.gonzalez14, eladio.quintero1, felisindo.garcia, antony.garcia, yessica.saez, edwin.collado}@utp.ac.pa

² Centro de Estudios Multidisciplinarios en Ciencias, Ingeniería y Tecnología AIP (CEMCIT AIP, Panamá

*Corresponding author: edwin.collado@utp.ac.pa

Resumen – *La creciente preocupación por la contaminación del aire y el cambio climático, derivada del aumento industrial global, ha generado perturbaciones ambientales de consecuencias imprevisibles para la salud humana. Entre los agentes contaminantes más dañinos se encuentran las partículas suspendidas en el aire, conocidas como material particulado, cuya dispersión rápida en áreas extensas agrava el problema. Con el objetivo de abordar esta problemática, se desarrolló un sistema de monitoreo atmosférico urbano basado en el Internet de las cosas que permite medir las concentraciones de material particulado PM2.5 y PM10. Este sistema está compuesto principalmente por un sensor Nova PM SDS011 para tomar mediciones de PM10 y PM2.5, un sensor DHT22 para registrar los niveles de temperatura y humedad y microcontrolador ESP32 programado en MicroPython, el cual se encargó de la recolección y transmisión de datos vía WiFi. El sistema fue validado utilizando el AEROQUAL S500 como equipo de referencia, donde se llegó a la conclusión que la estación de medición propuesta mide satisfactoriamente las variables estudiadas. Sin embargo, durante el periodo de validación se pudo observar que parte de la data recolectada presentaba errores, ya sea porque la medición fue corrompida, el sistema no funcionó correctamente y/o se registró un valor no deseado. Por ello, en este trabajo se propone desarrollar una herramienta basada en Ciencia de Datos para procesar adecuadamente los datos y reducir la cantidad de errores en las mediciones recolectadas. Específicamente, esta herramienta utiliza Python para aplicar técnica de limpieza de valores atípicos basada en el Rango Inter cuartilico. Esto permite identificar y corregir lecturas erróneas causadas por factores externos. Con esto se busca generar información precisa sobre los niveles de contaminación del aire para apoyar la toma de decisiones. El prototipo desarrollado demuestra que, mediante el uso de Internet de las Cosas y técnicas sencillas de Ciencia de Datos, es posible brindar un sistema de monitoreo atmosférico urbano de alta precisión, bajo costo y datos fiables, permitiendo una mejor comprensión de los riesgos a la salud para apoyar la adopción de medidas de protección ambiental.*

Palabras claves-- *Contaminación atmosférica, ciencia de datos, material particulado, métodos estadísticos, sensores.*

I. INTRODUCCIÓN

En los últimos años, la humanidad ha experimentado notables avances tecnológicos e industriales que, si bien han aportado beneficios significativos, también han desencadenado un consumo masivo de recursos naturales y energéticos. Esta voraz demanda ha generado consecuencias perjudiciales, manifestándose en forma de considerables desechos y emisiones contaminantes que impactan directamente en la atmósfera.

La contaminación atmosférica, también conocida como contaminación del aire, se ha convertido en un problema crítico para la salud humana. Proteger a las poblaciones de la mala calidad del aire es crucial y un desafío que concierne a todas las naciones del mundo [1]. Los efectos dañinos de estos contaminantes van desde la muerte, problemas cardiovasculares, enfermedades pulmonares e infecciones respiratorias agudas, entre otras afectaciones. Según estimaciones de la OMS (Organización Mundial de la Salud), la principal causa de estos males se debe a la gran exposición de material particulado. Esta exposición está directamente relacionada con la morbilidad y mortalidad tanto a corto como a largo plazo. Incluso en niveles considerados bajos, pueden dañar el medio ambiente, reducir la visibilidad y dañar materiales y estructuras [2]. Estas partículas se generan en la atmósfera como resultado de complejas reacciones químicas y se clasifican principalmente en dos categorías de partículas suspendidas respirables: PM10 (partículas con un diámetro promedio de 10 micrómetros) y PM2.5 (partículas con un diámetro promedio de 2.5 micrómetros) [3].

La preocupación por la contaminación atmosférica ha llevado a la OMS a actualizar sus directrices de calidad del aire en 2021, con el objetivo de evaluar los niveles perjudiciales para la salud y orientar los esfuerzos hacia la reducción de emisiones contaminantes [2], [4]. Según estas nuevas directrices, se establecen límites máximos permisibles para diversos contaminantes atmosféricos, incluido el material particulado. Sin embargo, en muchos países, incluyendo Panamá, la ausencia de políticas nacionales específicas

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

conlleva a la adopción de estas directrices de la OMS como referencia [5].

La problemática vinculada a la contaminación por material particulado se erige como una cuestión de relevancia trascendental en la actualidad. La imperiosa necesidad de reducir su presencia en la atmósfera y mejorar la calidad del aire no puede sino ser abordada de manera global, requiriendo un esfuerzo conjunto de todas las sociedades. La obtención de un conocimiento preciso acerca de la concentración de partículas en el aire de una ciudad específica adquiere un papel fundamental, ya que posibilita una conciencia más aguda sobre los niveles de contaminación presentes, permitiendo la implementación de medidas efectivas para mitigar sus impactos negativos.

La creciente preocupación por la contaminación atmosférica ha llevado a la comunidad científica a intensificar sus esfuerzos en la investigación y desarrollo de soluciones tecnológicas para medir la concentración de material particulado. El uso de sensores de bajo costo puede llevar una vía eficaz para llevar a cabo las mediciones de contaminación atmosférica, pero todos estos dispositivos de medición tienen cierta incertidumbre asociada, incluso en pruebas de laboratorio. Por lo tanto, la detección de valores atípicos y el uso de herramienta para corregir estos errores en estaciones de monitoreo es fundamental para aumentar la fiabilidad en la medición [1], [6]. En el estudio realizado en [7] los autores demuestran que los sistemas de sensores de calidad del aire de bajo costo pueden caracterizar adecuadamente las concentraciones de contaminación ambiental urbana en escalas de tiempo rápidas. En [8], discuten la evaluación del rendimiento de dispositivos de medición, específicamente de sensores de bajo costo en condiciones de laboratorio y de campo y destacan la importancia de una evaluación cuidadosa del rendimiento de los dispositivos y la necesidad de considerar múltiples factores al interpretar los resultados de estudios en su colocación. La consideración de dichos factores es crucial y así lo demuestra la investigación realizada por [6], donde desarrollaron correcciones específicas a sensores de bajos costo de material particulado para mejorar la precisión de las mediciones, especialmente teniendo en cuenta los efectos de la temperatura y la humedad permitiendo obtener información más precisa y mejor comprensión de la calidad del aire. Por su parte, en el estudio presentado en [9], utilizaron técnicas de regresión múltiple de datos a concentraciones de cuatro contaminantes atmosféricos (SO₂, NO₂, H₂S y PM₁₀) y utilizaron un análisis de regresión lineal múltiple paso a paso en la selección de variables regresores, donde las más importantes fueron la temperatura, la humedad relativa y la dirección del viento. Se obtuvieron funciones de regresión para la concentración anual, mensual y diaria de estos contaminantes. Se concluyó que las variables meteorológicas describieron adecuadamente la concentración anual y mensual, pero no la concentración diaria de los contaminantes. En [10], se demuestra que el modelo de *Random Forest* (RF) aplicado al

paquete de sensores de bajo costo, puede caracterizar con precisión las concentraciones de contaminantes atmosféricos en áreas urbanas de Estados Unidos y Europa. Los modelos RF lograron errores fraccionales bajos para CO₂, CO, O₃ y NO₂, cumpliendo con las métricas de precisión y exactitud recomendadas. Por otra parte, en [11] se presenta una metodología efectiva para mejorar la precisión de las mediciones de PM_{2,5} realizadas por el sensor PurpleAir, mediante la limpieza de datos y la corrección de sesgos. Se desarrolló un modelo de corrección estadounidense único que incorpora datos de PM_{2,5} y humedad relativa, lo que redujo significativamente el error en las mediciones y mejoró la precisión. En el trabajo desarrollado en [12] realizan una evaluación exhaustiva de varios algoritmos de corrección de datos para sensores de gases electroquímicos (ECGS) utilizando un conjunto de datos de comparaciones de campo y se dependencia del sesgo con la temperatura, la humedad relativa, los niveles de gas objetivo y la sensibilidad cruzada mediante diferentes algoritmos de corrección. Se proporcionaron recomendaciones sobre la selección del esquema de corrección de datos en función de los resultados de la comparación. A su vez, los autores en [13] presentan un nuevo método de detección de valores atípicos llamado VGOD, diseñado específicamente para redes de sensores de contaminación del aire. El proceso de detección consta de tres etapas: aprendizaje de un gráfico, reconstrucción de señales y la inspección de los residuos, donde se sugiere explorar la aplicabilidad de redes neuronales gráficas para mejorar la detección de valores atípicos en este tipo de redes de sensores.

En el trabajo realizado por [14] se utilizó el método de detección de valores atípicos del diagrama de caja para identificar anomalías en la precisión y exactitud de sistemas de sensores de bajo costo de medición de partículas PM_{2.5} en el aire. Asimismo, estudios epidemiológicos han demostrado asociaciones entre el aumento en el rango intercuartílico de PM_{2.5} y efectos adversos cardiopulmonares [15],[16]. También se menciona la importancia de seleccionar un tamaño adecuado de cuadrícula para la detección de anomalías en el método [14],[16].

Con base en los estudios antes mencionados, el objetivo de este trabajo es desarrollar un sistema de monitoreo atmosférico portátil, eficiente y asequible basado en el Internet de las Cosas (IoT) y Ciencia de Datos para medir, analizar y procesar los datos de las concentraciones de material particulado PM_{2.5} y PM₁₀ en áreas urbanas. Esta herramienta se implementa utilizando Python para aplicar una técnica de limpieza de datos basada en el Rango Intercuartílico (IQR) con la idea de mejorar la precisión de las mediciones. El sistema de monitoreo de contaminación atmosférica fue propuesto en el trabajo presentado en [17], que está compuesto por un sensor Nova PM SDS011 medir material particulado, un sensor DHT22 para registrar los niveles de temperatura y humedad relativa y microcontrolador ESP32 programado en MicroPython, el cual se encargó de la recolección y transmisión de datos vía WiFi a

una plataforma de visualización de datos en línea para un análisis en tiempo real.

Con esta herramienta se busca proporcionar a entidades gubernamentales, empresas y la población en general información precisa sobre la calidad del aire y su impacto en la salud y el medio ambiente, lo que puede influir en la toma de decisiones para mejorar la calidad de vida y el bienestar de los ciudadanos.

El artículo está organizado de la siguiente manera: La sección II describe el diseño y metodología utilizada en el proyecto. La sección III los resultados y discusión. La sección IV presenta las conclusiones.

II. DISEÑO Y METODOLOGÍA

Este proyecto tiene como objetivo desarrollar una herramienta basada en Ciencias de Datos, específicamente una técnica de limpieza de datos basada en el Rango Inter cuartílico (IQR), para procesar adecuadamente los datos y reducir la cantidad de errores en las mediciones recolectadas en los sistemas de monitoreo de partículas suspendidas PM10 y PM2.5. En esta sección, se detallará la formulación del problema, el modelo conceptual, los materiales utilizados, el método y la técnica utilizada, la ubicación y la recopilación de datos. Con este proyecto se busca crear una herramienta integrada con tecnología accesible y actualizada capaz de monitorear la calidad del aire y brindar datos confiables para contribuir con la toma de decisiones informadas y la adopción de medidas preventivas para mejorar la salud y el medio ambiente.

A. Formulación del problema

La calidad del aire es un tema crítico para la salud y el bienestar de la población. El material particulado (PM) es uno de los contaminantes del aire más peligrosos, ya que puede penetrar profundamente en los pulmones y el sistema cardiovascular y causar enfermedades respiratorias, cardiovasculares y otros problemas de salud graves. Las concentraciones de PM pueden variar considerablemente de un lugar a otro y de un momento a otro, lo que dificulta la evaluación precisa de los riesgos para la salud y la toma de decisiones informadas. Ante esta problemática, en el trabajo presentado en [17], se desarrolló un prototipo de sistema de bajo costo y basado en IoT para el monitoreo atmosférico urbano en interiores y exteriores. Este sistema está compuesto por un sensor de material particulado Nova PM SDS011 para medir partículas como el polvo, la suciedad, el hollín y/o el humo en el aire. Además, cuenta con un sensor DHT22 para medir temperatura y humedad relativa, variables que pueden afectar el comportamiento de los contaminantes. Estos sensores están conectados mediante una placa PCB al microcontrolador ESP32, que se encarga de procesar los datos del sensor, realizar tareas de monitoreo y enviar la información a través de una red de comunicación inalámbrica a Internet a la base de datos de la plataforma de visualización de datos ThingSpeak. Esta

plataforma permite al usuario analizar los niveles de contaminación en tiempo real mediante gráficos y tablas, permitiendo también exportar esta información para su uso en otros equipos con programas especializados. La Fig. 1 muestra el modelo conceptual del prototipo.

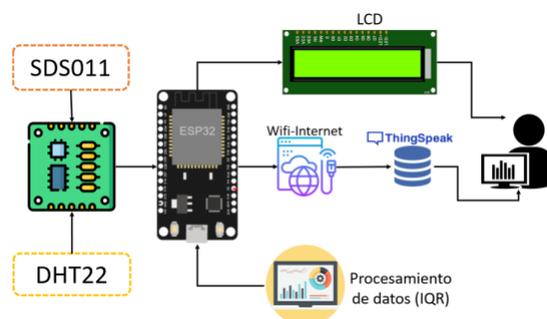


Fig. 1 Modelo conceptual del sistema de monitoreo de PM.

Durante el periodo de validación se pudo observar que parte de la data recolectada presentaba errores, ya sea porque la medición fue corrompida, el sistema no funcionó correctamente y/o se registró un valor no deseado. Por ello, en este trabajo se propone desarrollar una herramienta basada en Ciencia de Datos para procesar adecuadamente los datos y reducir la cantidad de errores en las mediciones recolectadas.

B. Propuesta de herramienta basada en Ciencia de Datos para reducir errores en las mediciones en sistemas de monitoreo atmosférico

Para solventar el problema antes mencionado, este trabajo propone desarrollar una herramienta basada en Ciencia de Datos, específicamente una técnica de limpieza de datos basada en el Rango Inter cuartílico (IQR), para mejorar el proceso de recolección de datos y reducir la cantidad de errores en las mediciones. El sistema de monitoreo de material particulado con la herramienta para reducción de errores de medición fue basado en el trabajo realizado por los autores en [13]. Esta herramienta fue implementada en el microcontrolador ESP32 utilizando un lenguaje Micro-Python para efectuar lecturas periódicas de los sensores, aplicar algoritmos de limpieza de valores de medición atípicos y empaquetar la información para transmisión utilizando el protocolo MQTT a través de WiFi.

El método del Rango Inter cuartílico (IQR) es una herramienta estadística utilizada para identificar valores atípicos en un conjunto de datos, esta es una medida de dispersión estadística que describe la variabilidad en un conjunto de datos y se utiliza comúnmente para identificar valores atípicos. Se calcula como la diferencia entre el tercer cuartil y el primer cuartil de los datos y es una técnica robusta que no se ve afectada por los valores extremos. Por estas razones, el IQR fue elegido el método principal para la limpieza de datos de PM en este proyecto.

La expresión matemática para medir el IQR es:

$$IQR = Q_3 - Q_1 \quad (1)$$

El IQR es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1) de un conjunto de datos. Representa el rango medio de los datos, excluyendo los valores extremos. Los cuartiles son los puntos que dividen un conjunto de datos en cuatro partes iguales. Q1 es el valor medio entre el número más pequeño y la mediana del conjunto de datos, mientras que Q3 es el valor medio entre la mediana y el valor más alto [18].

Los valores atípicos se definen como aquellos que están por debajo de $Q_1 - 1.5 * IQR$ o por encima de $Q_1 + 1.5 * IQR$. Estos valores se consideran atípicos y pueden ser eliminados del conjunto de datos.

La implementación de la limpieza de valores atípicos se realizó utilizando Python y se siguió el siguiente enfoque:

1. Se calcularon los valores de Q1 y Q3 para las distribuciones de concentraciones de PM10 y PM2.5.
2. Se calculó el IQR para cada conjunto de datos.
3. Se identificaron los valores atípicos utilizando los criterios mencionados anteriormente.
4. Se limitaron los valores atípicos del conjunto de datos

Para tratar los valores atípicos identificados mediante los límites superior e inferior, se optó por la técnica de “truncamiento” o limitar estos valores. Esto significa que todos los valores atípicos identificados se ajustan al límite más cercano, ya sea el superior o inferior, dependiendo de si son extremadamente altos o bajos. Este método preserva todos los datos, evitando la pérdida de filas por eliminación y reduce el impacto de los extremos sin alterar significativamente la estructura general de los datos.

El código implementado es el siguiente:

Algoritmo 1 Corrección de Valores Atípicos

```

1: procedure CORREGIRVA(df, cols) ▷ Corrige los valores
   atípicos en un DataFrame
2: for col in cols do ▷ Itera sobre las columnas
   especificadas
3:   Q1 ← df[col].quantile(0,25) ▷ Primer cuartil
4:   Q3 ← df[columna].quantile(0,75) ▷ Tercer cuartil
5:   IQR ← Q3 - Q1 ▷ Rango intercuartílico
6:   Linf ← Q1 - 1.5 × IQR ▷ Límite inferior
7:   Lsup ← Q3 + 1.5 × IQR ▷ Límite superior
8:   df [col] ← df [col].clip(Linf, Lsup) ▷ Ajusta valores fuera
   de límites
9: end for
10: return df ▷ Devuelve el DataFrame corregido
11: end procedure

```

Como se mencionó anteriormente, el IQR fue implementado en un entorno de programación Python. Se utilizó BoxPlot junto con histogramas para visualizar la distribución cuantitativa de los datos y el efecto de la limpieza de valores atípicos. Sin embargo, es importante destacar que, si

bien esta técnica es efectiva y ampliamente aceptada, existe la posibilidad de eliminar información valiosa, como eventos extraordinarios que puedan haber ocurrido durante el período de muestreo.

III. RESULTADOS Y DISCUSIÓN

El sistema de monitoreo de material particulado está ubicado en el distrito de Juan Diaz, Ciudad de Panamá, debido a que este punto cuenta con una alta densidad de población para realizar mediciones y de esta forma tener datos representativos del nivel de contaminación de aire en el que están expuestas las personas. Para evaluar la concentración de material particulado, se llevaron a cabo mediciones de PM2.5, PM10, temperatura y humedad relativa durante un 1 año con registros cada 5 minutos, lo que resultó en un total de 248, 882 lecturas. A continuación, se procederá al análisis de estos datos.

Los resultados obtenidos durante las pruebas en el entorno mencionado previamente han sido alentadores, confirmando la efectividad del prototipo desarrollado. En primer lugar, se ha validado la precisión del sensor SDS011 al compararlo con un sensor comercial de alta precisión AEROQUAL S500, según los estudios realizados por el grupo de investigación ITSIAS [17], [19]. Estos resultados han sido prometedores, mostrando un mínimo margen de error porcentual, lo que respalda la confiabilidad del sensor Nova PM SDS011 para llevar a cabo mediciones precisas. Para analizar el comportamiento de la contaminación por material particulado, los datos recopilados por el prototipo se exportaron a una computadora y se procesaron utilizando el entorno de programación Python. En la Fig. 2 se muestran los niveles de concentración de material particulado PM10 y PM2.5 antes de implementar la herramienta para reducir los valores atípicos.

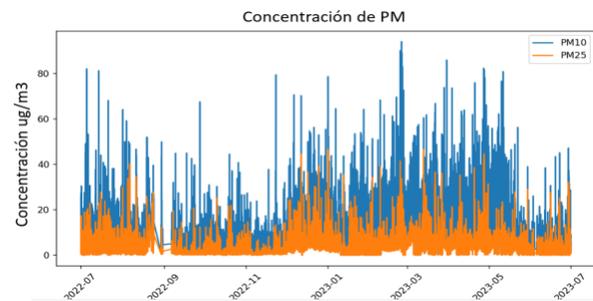


Fig. 2 Datos de concentración de PM₁₀ y PM_{2.5} inicialmente.

Para visualizar el efecto de la limpieza de valores atípicos, se generaron histogramas y BoxPlot, proporcionando una representación gráfica de la distribución de las concentraciones de PM10 y PM2.5 antes y después de la limpieza de datos. Las Fig. 3 y 4 muestran el histograma y BoxPlot de las concentraciones de PM10 y PM2.5 antes de implementar el IQR, donde se pueden observar la presencia de valores atípicos.

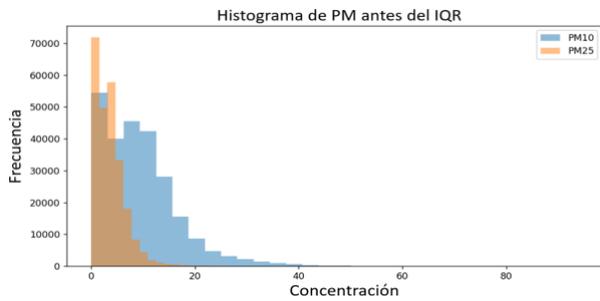


Fig. 3 Histograma de material particulado antes de la corrección de datos mediante IQR.

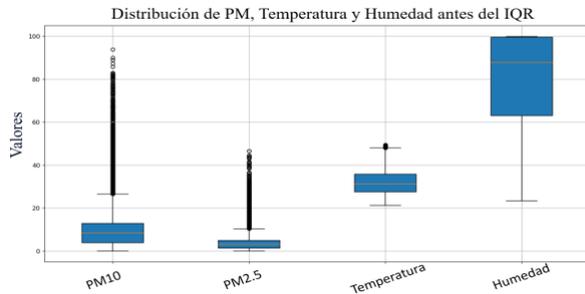


Fig. 4 BoxPlot de PM, temperatura y humedad antes de la corrección de datos mediante IQR.

Como se puede apreciar en la Fig. 4 se muestra una gran cantidad de valores atípicos que se encuentran significativamente alejados del cuerpo principal de la distribución. Esto se puede apreciar más en el histograma de la Fig. 3 donde ciertos datos mayores a $40 \mu\text{g}/\text{m}^3$ presentan una frecuencia muy mínima producto de errores en la medición por el sensor u otras interferencias.

Las Fig. 5 y 6 muestran el histograma y BoxPlot de las concentraciones de PM10 y PM2.5 después de implementar el IQR, para reducir la presencia de valores atípicos.

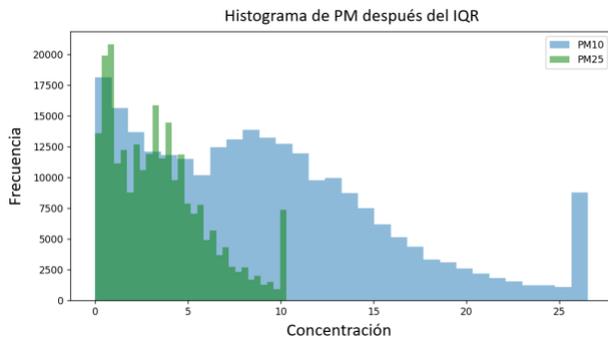


Fig. 5 Histograma de material particulado después de la corrección de datos mediante IQR.

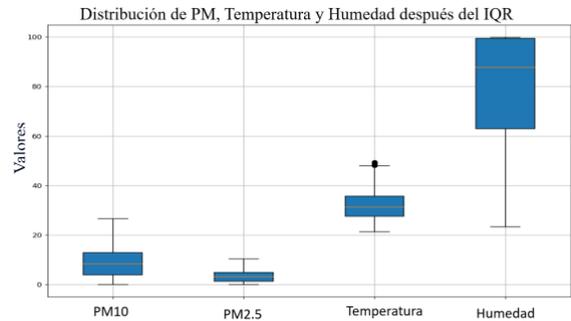


Fig. 6 BoxPlot de PM, temperatura y humedad después de la corrección de datos mediante IQR.

En la Fig. 7 se muestran los niveles de concentración de material particulado PM10 y PM2.5 después de implementar la herramienta para reducir los valores atípicos.

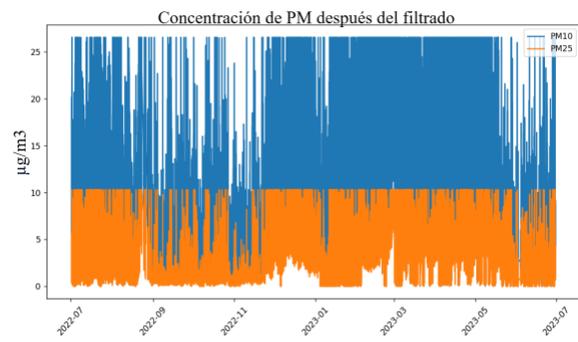


Fig. 7 Datos de concentración de PM después del filtrado.

La Fig. 6 y 7 muestran que gracias a la corrección de datos mediante la técnica IQR se puede ver una gráfica para los datos de PM sin la presencia de errores.

Después de implementar el IQR para reducir los errores en las mediciones, se puede observar un aumento en los niveles de contaminación en momentos específicos del día, lo que sugiere la presencia de factores desencadenantes. Específicamente, se ha notado un incremento en la contaminación por PM entre las 5:00 y las 8:00 a.m., y entre las 4:00 y las 6:00 p.m., indicando que el uso de vehículos por parte de la población podría ser un factor determinante.

Además, se han registrado los valores máximo, mínimo y promedio de concentración, los cuales son de gran interés para el análisis y comprensión. Para los datos de PM10 se registró una concentración mínima de $0.6 \mu\text{g}/\text{m}^3$ un máximo de $93.4 \mu\text{g}/\text{m}^3$ y un promedio de $9.15 \mu\text{g}/\text{m}^3$. Para los datos de PM2.5 se registró una concentración mínima de $0.5 \mu\text{g}/\text{m}^3$ un máximo de $47 \mu\text{g}/\text{m}^3$ y un promedio de $3.55 \mu\text{g}/\text{m}^3$.

En las Figuras 8 y 9 se muestran la variación de temperatura y humedad relativa durante el periodo de estudio, en las cuales se observa que a medida incrementa la temperatura, la humedad disminuye proporcionalmente. El análisis del comportamiento de estas variables es de mucha utilidad para entender el impacto de variables climáticas en los contaminantes de PM.

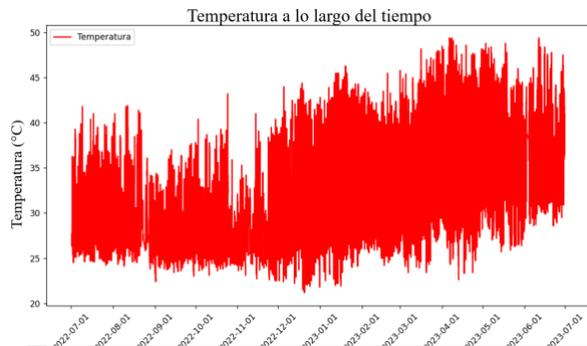


Fig. 8 Datos del comportamiento de la temperatura.

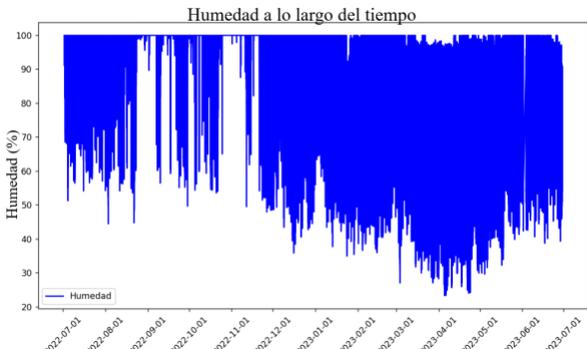


Fig. 9 Datos del comportamiento de la humedad relativa.

IV. CONCLUSIONES

La creciente problemática de la contaminación del aire y el cambio climático presenta un desafío considerable para la salud pública y el medio ambiente, siendo impulsada por el constante crecimiento industrial y la expansión urbana. Entre los diversos contaminantes, el material particulado se destaca como uno de los más perjudiciales para la salud humana debido a sus efectos adversos potenciales y su capacidad para dispersarse rápidamente en áreas extensas.

El desarrollo y las pruebas de validación realizadas al prototipo de sistema de monitoreo atmosférico basado en IoT demuestran que es posible implementar un sistema de medición de material particulado de bajo costo y alta eficiencia utilizando componentes accesibles. La incorporación del sensor Nova PM SDS011 proporcionó lecturas precisas de PM10 y PM2.5, mientras que el microcontrolador ESP32 gestionó eficientemente la recolección y transmisión de datos a la nube.

La visualización con histogramas y BoxPlot ha sido fundamental para comprender la distribución cuantitativa de los datos antes y después del proceso de limpieza. Estas herramientas han proporcionado una representación gráfica clara de los efectos de la limpieza de valores atípicos en la distribución de las concentraciones de PM10 y PM2.5, permitiendo una evaluación visual del éxito de la técnica aplicada.

La técnica de limpieza de datos basada en el Rango Intercuartílico (IQR) implementada en Python resultó efectiva para identificar y corregir valores atípicos en los datos de PM.

Esto mejoró significativamente la precisión de las mediciones al eliminar lecturas erróneas causadas por interferencias temporales.

El análisis de los datos recolectados reveló patrones importantes sobre los niveles y fluctuaciones de PM a lo largo del día, indicando que el tráfico vehicular en horas pico podría ser un factor determinante. Asimismo, se observó la relación inversa entre temperatura y humedad.

Este proyecto demuestra que los sistemas de monitoreo atmosférico basados en IoT, implementando técnicas sencillas de ciencia de datos, pueden generar información precisa sobre la calidad del aire urbano. Esto permite una mejor comprensión de los riesgos a la salud y apoya la adopción de medidas efectivas por parte de las autoridades para proteger a la población. Como trabajo futuro se espera implementar el sistema en varios puntos de medición simultáneamente para crear una red más amplia de nodos de sensores.

AGRADECIMIENTO

Agradecemos al Laboratorio de TIC para Monitoreo Ambiental del grupo de investigación ITSIAS – UTP y al Laboratorio de Sistemas Inteligentes (LABSI) por su apoyo durante el diseño y la fabricación del prototipo. Los autores agradecen también el apoyo de la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT), el Sistema Nacional de Investigación (SNI) por su apoyo en la realización de la investigación.

REFERENCIAS

- [1] Y. Quintero, Y. Rodríguez y A. Bello. "Vista de Efecto de la contaminación atmosférica y calidad del aire". *Revistas SENA*, 2020. [Online]. Available: <https://revistas.sena.edu.co/index.php/riag/article/view/4655/4887>
- [2] "Ambient air pollution (outdoors)". WHO | World Health Organization, 2022. [Online]. Available: [https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- [3] A. Rojas Bardales y M. Guerra-Saldaña, "Nivel de contaminación del aire urbano por partículas suspendidas respirables (PM-10 y PM-2,5)", *Rev. Amaz. Cienc. Ambient. Ecol.*, vol. 1, núm. 2, p. e373, 2022.
- [4] Salud sin Daño. "Guías actualizadas de la OMS sobre la calidad del aire y sus implicancias para los países latinoamericanos". Health Care Without Harm, 2022. [Online]. Available: <https://saludsindanio.org/documentos/guia-calidad-aire-OMS>
- [5] F. P. Verdezoto Mendoza, J. C. Muyulema Allaica, H. R. Cuba Torre, y A. K. Serrano Castro, "Evaluación temporal del material particulado PM2.5 y PM10 en el Distrito Metropolitano de Quito – Ecuador", *ConcienciaDigital*, vol. 5, núm. 4.1, pp. 21–44, 2022.
- [6] C. Malings *et al.*, "Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation", *Aerosol Sci. Technol.*, vol. 54, núm. 2, pp. 160–174, 2020.
- [7] E. S. Cross *et al.*, "Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements", *Atmos. Meas. Tech.*, vol. 10, núm. 9, pp. 3575–3588, 2017.
- [8] S. Diez *et al.*, "Air pollution measurement errors: is your data fit for purpose?", *Atmos. Meas. Tech.*, vol. 15, núm. 13, pp. 4091–4105, 2022.
- [9] E. E. de La Correlación Entre Contaminantes Atmosféricos Y Variables Meteorológicas En La Zona Norte de Chiapas, "Statistical study of the correlation between atmospheric pollutants and meteorological", *Org.mx*. [Online]. Available: <https://www.scielo.org.mx/pdf/uc/v26n1/v26n1a5.pdf>

- [10] N. Zimmerman *et al.*, "A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring", *Atmos. Meas. Tech.*, vol. 11, núm. 1, pp. 291–313, 2018.
- [11] K. K. Barkjohn, B. Gantt, y A. L. Clements, "Development and application of a United States-wide correction for PM_{2.5} data collected with the PurpleAir sensor", *Atmos. Meas. Tech.*, vol. 14, núm. 6, pp. 4617–4637, 2021.
- [12] Y. Liang *et al.*, "Field comparison of electrochemical gas sensor data correction algorithms for ambient air measurements", *Sens. Actuators B Chem.*, vol. 327, núm. 128897, p. 128897, 2021.
- [13] P. Ferrer-Cid, J. M. Barcelo-Ordinas and J. Garcia-Vidal, "Volterra Graph-Based Outlier Detection for Air Pollution Sensor Networks," in *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2759–2771, 1 July-Aug. 2022.
- [14] C.-J. Liang y P.-R. Yu, "Assessment and improvement of two low-cost particulate matter sensor systems by using spatial interpolation data from air quality monitoring stations", *Atmosphere (Basel)*, vol. 12, núm. 3, p. 300, 2021.
- [15] J. Cao, H. Xu, Q. Xu, B. Chen, y H. Kan, "Fine particulate matter constituents and cardiopulmonary mortality in a heavily polluted Chinese city", *Environ. Health Perspect.*, vol. 120, núm. 3, pp. 373–378, 2012.
- [16] Y. Tian *et al.*, "Short-term effects of ambient fine particulate matter pollution on hospital visits for chronic obstructive pulmonary disease in Beijing, China", *Environ. Health*, vol. 17, núm. 1, 2018.
- [17] Quintero E., González J., Garcia F., Sáez Y., and Collado, E.. "IoT-based system prototype for particulate matter monitoring in the city of Chitre, Panama". 21st LACCEI International Multi-Conference for Engineering, Education, and Technology. Buenos Aires, Argentina. July 17 - 21, 2023.
- [18] A. Latam, "Mejora del análisis con Boxplot", *Alura*, 24-mar-2023. [Online]. Available: <https://www.aluracursos.com/blog/mejora-del-analisis-con-boxplot>
- [19] De León O., Cedeño, B., Collado E., Sáez, Y., and Canto F. "Diseño de un Sistema basado en Internet de las Cosas para Monitoreo de Contaminantes de Aire". 19th LACCEI International Multi-Conference for Engineering, Education, and Technology. Bogota, DC, Colombia. July 19 - 23, 2021.