



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO®
Campus Nogales



PROTOTIPO DE UNA RED MÍNIMA DE MONITOREO DE CONTAMINANTES EN AMBIENTES CERRADOS

TESIS

QUE PARA OBTENER EL GRADO DE
MAESTRO EN SISTEMAS COMPUTACIONALES

PRESENTA

BRAYAN ISAÍ PÉREZ VALENZUELA

DIRECTOR

M.C. SIGIFREDO GARCÍA ALVA

H. NOGALES, SONORA, MÉXICO.

JUNIO DE 2023.

Nogales, Sonora, **21/junio/2023**
Oficio No. DEPI/180/2023

MANUEL OMAR MERANZA CASTILLON
DOCENTE DEL TECNM/CAMPUS NOGALES
PRESENTE

Por este medio le comunicamos a Usted que el trabajo de Tesis denominado: "PROTOTIPO DE UNA RED MÍNIMA DE MONITOREO DE CONTAMINANTES EN AMBIENTES CERRADOS", que presentó el alumno BRAYAN ISAÍ PÉREZ VALENZUELA, con número de control 19349003, candidato a obtener el grado de MAESTRO EN SISTEMAS COMPUTACIONALES, ha sido revisado por los miembros del Comité Tutorial y cubiertas las observaciones realizadas, se Autoriza su Impresión y se Acepta para su Evaluación en la presentación del Examen de Grado.

Sin más por el momento, le reitero mi consideración distinguida.

ATENTAMENTE

Excelencia en Educación Tecnológica
"La Ciencia y la Tecnología para la Liberación del Hombre"



INSTITUTO TECNOLÓGICO DE NOGALES
DIVISIÓN DE ESTUDIOS DE,
POSGRADO E INVESTIGACIÓN

GUILLERMINA MUÑOZ ZAMORA
REVISORA

JESUS RAÚL CRUZ RENTERÍA
REVISOR

SIGIFREDO GARCÍA ALVA
REVISOR





Nogales, Sonora, **21/junio/2023**
Oficio No. DEPI/181/2023

BRAYAN ISAÍ PÉREZ VALENZUELA
CANDIDATO A OBTENER EL GRADO DE MAESTRO EN SISTEMAS COMPUTACIONALES
PRESENTE

De acuerdo con el Reglamento de Titulación del Sistema Nacional de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que el Comité Tutorial realizó, con respecto a su Tesis titulada: "PROTOTIPO DE UNA RED MÍNIMA DE MONITOREO DE CONTAMINANTES EN AMBIENTES CERRADOS", la División de Estudios de Posgrado e Investigación Autoriza su Impresión.

Sin más por el momento, le reitero mi consideración distinguida.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"La Ciencia y la Tecnología para la Liberación del Hombre"®

MANUEL OMAR MERANZA CASTILLON
JEFE DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN



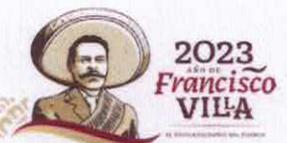
SEP
SECRETARÍA
DE EDUCACIÓN
PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE NOGALES
DIVISIÓN DE ESTUDIOS DE,
POSGRADO E INVESTIGACIÓN

ccp. Archivo



**PROTOTIPO DE UNA RED MÍNIMA DE
MONITOREO CONTAMINANTES EN
AMBIENTES CERRADOS**

Resumen

El presente documento se compone de 4 fases, en la primera de estas, se explican los requerimientos necesarios para diseñar el prototipo y el circuito de la red de monitoreo de contaminantes, así mismo se determina el tipo de sensores que se utilizarán para la detección de los contaminantes en el aire interior.

En la segunda fase se identifican los modelos matemáticos para aplicar el ajuste necesario a los sensores de gases contaminantes, así como los requerimientos necesarios para la programación del hardware utilizado, permitiendo evaluar el comportamiento y comparar el funcionamiento del prototipo al exponerlo a concentraciones reales de las partículas de gases contaminantes.

La tercera fase explica las características iniciales de los datos que se utilizaron y el proceso necesario, para que se combinaran en un solo formato apropiado para llevar a cabo la investigación.

En la cuarta etapa, se lleva a cabo la implementación de un análisis multivariado que engloba varios algoritmos de aprendizaje profundo. Los resultados obtenidos indican que las sustancias detectadas en el aire interior no presentan relaciones entre sí y su variabilidad está principalmente condicionada por los factores climáticos.

Abstract

This document is made up of 4 phases, in the first of which the necessary requirements are explained to design the prototype and the circuit of the contaminant monitoring network, likewise the type of sensors that will be used for the detection of contaminants is determined. in indoor air.

In the second phase, the mathematical models are identified to apply the necessary adjustment to the polluting gas sensors, as well as the necessary requirements for the programming of the hardware used, allowing to evaluate the behavior and compare the operation of the prototype when exposed to real concentrations of particles. of polluting gases.

The third phase explains the initial characteristics of the data that were used and the process that was needed to combine them into a single appropriate format to carry out the investigation.

In the fourth stage, the implementation of a multivariate analysis that includes several deep learning algorithms is carried out. The results obtained indicate that the substances detected in indoor air have no relationship with each other, and their preference is mainly conditioned by climatic factors.

Dedicatoria

A mis padres, por todo lo que hicieron por mí a lo largo de mi vida. Su dedicación, amor incondicional y sacrificio han dejado una huella imborrable en mi corazón, y quisiera que supieran cuánto valoro cada momento que compartimos juntos. Su guía y apoyo han sido fundamentales en mi crecimiento y desarrollo. Desde los pequeños gestos cotidianos, hasta los grandes logros, siempre sentí su respaldo y aliento. Sus palabras de sabiduría y su paciencia infinita me ayudaron a superar obstáculos y a convertirme en la persona que soy hoy.

En cada paso de mi camino, siempre sentiré su amor incondicional como un faro que me guía. Aunque sé que no podrán leer estas palabras nunca me cansaría de expresar completamente la gratitud que siento, ya que su influencia positiva me dio la fortaleza para enfrentar desafíos y la confianza para perseguir mis sueños.

Papá y Mamá, su amor ha sido mi refugio en los momentos más difíciles y mi fuente de alegría en los momentos felices. Su bondad, paciencia y amor eterno han dejado una marca profunda en mi corazón y en mi vida. Hoy quiero tomar un momento para decir simplemente: ¡Gracias! Gracias por haber sido los padres increíbles que deseé tener y por todo lo que hicieron por mí.

Con todo mi amor y agradecimiento para ustedes.

Agradecimientos

A mi familia, quiero agradecerles por ser mi base, mi sostén y mi inspiración. Su amor incondicional, apoyo constante y sus consejos sabios han sido fundamentales en mi crecimiento y desarrollo. Cada uno de ustedes ha contribuido de manera única a hacer de mí una mejor persona y por eso les estaré eternamente agradecido.

A mis amigos, quiero expresar cuánto valoro su amistad. Su presencia en mi vida ha iluminado mis días y ha hecho que los momentos especiales sean aún más significativos. Sus risas, su apoyo y su compañerismo han sido un faro en los días oscuros y un motivo de alegría en los días felices.

Quiero agradecer a mi jefe el Ing. Gregorio Rodríguez Chavarín por las risas compartidas, por las experiencias inolvidables y por ser alguien en quien siempre puedo confiar. Tu presencia ha enriquecido mi vida de maneras que no podría haber imaginado y estoy profundamente agradecido por todo lo que has hecho por mí.

Quiero expresar mi más sincero agradecimiento a mi terapeuta físico Erick Sánchez por toda la dedicación, cuidado y profesionalismo que ha demostrado durante mi tratamiento. Su compromiso con mi bienestar y su habilidad para aliviar mis molestias han marcado una diferencia significativa en mi calidad de vida. Así como su enfoque comprensivo y su capacidad para adaptar las terapias a mis necesidades específicas fueron verdaderamente inspiradoras. Cada sesión con él ha sido un paso hacia la recuperación, estoy muy agradecido por el progreso que hemos logrado juntos. No solo es un excelente fisioterapeuta, sino también una persona cálida, amable y un buen amigo. Su apoyo y ánimo constante han hecho que, enfrentar los desafíos físicos, sea mucho más llevadero. Siempre me sentí escuchado y atendido, lo cual fue fundamental para mi proceso de rehabilitación. Quiero agradecerle por su paciencia en explicar cada ejercicio y técnica, por su empeño en mi bienestar y por brindarme las herramientas necesarias para continuar cuidándome incluso fuera de las sesiones. Su dedicación va más allá de lo que se espera y realmente valoro todo lo que ha hecho por mí. Gracias por ser un guía, un apoyo y un profesional excepcional en mi camino hacia la recuperación.

Agradecimientos institucionales

Agradezco al Instituto Tecnológico de Nogales por haberme abierto las puertas para continuar con mis estudios, así como también a los diferentes docentes que brindaron de sus grandes conocimientos y su apoyo para seguir adelante.

Quiero dar un agradecimiento especial a mi director de tesis M.C Sigifredo García Alva por haberme brindado todo su apoyo incondicional para que sobresaliera, a pesar de todas las adversidades a las que me enfrenté durante mis estudios de posgrado, gracias a sus enseñanzas, paciencia y consejos es que logro culminar con esta etapa.

Agradezco al M.C Jesús Raúl Cruz Rentería por todo su apoyo y consejos brindados durante estos años.

Índice General

CAPÍTULO I. ANTECEDENTES
1.1 Introducción	2
1.2 Estado del arte	5
1.2.1 Estimación de la calidad del aire dentro de diferentes laboratorios químicos.....	5
1.2.2 Prototipo de bajo costo para monitoreo del aire en ambientes cerrados	5
1.2.3 Sistema de sensores que detecta sustancias peligrosas en el aire	6
1.2.4 Uso de redes inalámbricas de sensores para monitorear gases contaminantes en espacios cerrados.....	6
1.2.5 Un sistema de monitoreo de seguridad de construcción en tiempo real para gases peligrosos que integra redes de sensores inalámbricos y tecnologías de modelado de información de edificios.....	7
1.3 Planteamiento del problema	8
1.4 Objetivos	9
1.4.1 Objetivo general.....	9
1.4.2 Objetivos específicos.....	9
1.5 Hipótesis.....	10
1.6 Justificación	11
1.6.1 Justificación social	11
1.6.2 Justificación tecnológica.....	11
1.6.3 Justificación económica.....	11
1.7 Variables, indicadores e instrumentos.....	12
1.7.1 Variable dependiente	12
1.7.2 Variables independientes.....	12
1.8 Limitaciones y Delimitaciones	13
1.8.1 Limitaciones.....	13
1.8.2 Delimitaciones.....	13
1.9 Metodología	14
CAPÍTULO II. MARCO TEÓRICO.....	..
2.1 Terminología utilizada para el estudio de la calidad del aire interior.....	16
2.1.1 Aire	16
2.1.2 Calidad del aire interior.....	16
2.1.3 Material particulado.....	16
2.1.4 Gases tóxicos	17

2.2	Técnicas utilizadas para el estudio de la contaminación del aire interior	17
2.2.1	Análisis multivariado	17
2.2.2	Diagrama de densidad.....	17
2.2.3	Aprendizaje automático	18
2.2.4	Estacionalidad	19
2.2.5	ARIMA.....	19
2.2.6	Series temporales.....	19
2.2.7	Prueba de Dickey Fuller aumentada	20
2.3	Términos y herramientas de software utilizadas.....	21
2.3.1	Datos	21
2.3.2	Base de datos	21
2.3.3	Excel Datastreamer	21
2.3.4	Minería de datos	22
2.3.5	Phyton	22
2.4	Componentes electrónicos	22
2.4.1	Sensores	23
2.4.2	Arduino	25
CAPÍTULO III. DESARROLLO
3.1	Fase de diseño y construcción	27
3.1.1	Arquitectura de la solución planteada	27
3.1.2	Arquitectura del sistema de sensores que incorpora el circuito	28
3.1.3	Arquitectura de almacenamiento de datos	28
3.1.4	Arquitectura de sistema electrónico.....	29
3.2	Definición de la estructura de los módulos de la red mínima de monitoreo de contaminantes de la calidad del aire interior	29
3.2.1	Requerimientos de la estructura.....	29
3.2.2	Diseño 2D de la estructura de los módulos de la red mínima de contaminantes de la calidad del aire interior	30
3.2.3	Vista 3D del prototipo armado.....	31
3.3	Fase de programación y pruebas	33
3.3.1	Arranque y ajuste preciso de los sensores.....	33
3.3.2	Calibración del sensor MQ-7	33
3.3.3	Calibración del sensor MQ135	35
3.3.4	Calibración del sensor de temperatura y humedad DHT11.....	36

3.3.5	Calibración del sensor SDS011 de partículas de PM_{10} y $PM_{2.5}$	36
3.3.6	Implementación y funcionamiento del prototipo.....	36
3.3.7	Funcionamiento del prototipo	37
3.4	Obtención, Interpretación y Extracción de Información.....	40
3.4.1	Procedencia de los datos.....	40
3.4.2	Obtención de los datos horarios de contaminación del aire interior	40
3.4.3	Descripción de los datos obtenidos	40
3.4.4	Pre procesado de los datos	41
3.4.5	Detección de los datos ausentes.....	42
3.4.6	Estandarización de los datos	43
3.4.7	Detección y eliminación de datos atípicos.....	44
3.4.8	Eliminación de datos atípicos.....	48
3.5	Fase de análisis exploratorio y visualización de los datos.....	51
3.5.1	Preparación de datos para representación gráfica	51
3.5.2	Exploración visual y extracción de la información	51
3.5.3	Estimación de la densidad del kernel (KDE)	58
3.5.4	Diagramas de conjunto o dispersión.....	61
3.5.5	Correlación lineal.....	63
3.5.6	Prueba de Dickey-Fuller aumentada.....	64
3.5.7	Árboles de decisión potenciados por gradientes.....	68
3.5.8	Correlación generada por los árboles de decisión	77
CAPÍTULO IV. RESULTADOS OBTENIDOS.....		.
4.1	Funcionamiento del prototipo	79
4.2	Análisis inicial de los datos.....	79
4.3	Análisis de Auto-regresión con modelo ARIMA	79
4.4	Resultados datos reales vs ARIMA	81
4.5	Resultados del análisis de correlación entre variables	83
4.6	Resultados de las predicciones de árboles de decisión	83
CAPÍTULO V. CONCLUSIONES.....		.
5.1	Conclusiones.....	85
5.2	Trabajos futuros	87
6	Referencias.....	88

Índice de figuras

Figura 2.1 Sensor MQ-7	23
Figura 2.2 Sensor de gases tóxicos	23
Figura 2.3 Sensor de partículas (Nova PM Sensor SDS011)	24
Figura 2.4 Sensor de humedad y temperatura	24
Figura 3.1 Arquitectura del sistema de monitoreo de calidad del aire en ambientes cerrados. Elaboración propia.....	27
Figura 3.2 vista isométrica del diseño del prototipo de los módulos de la red de contaminantes del aire interior. Elaboración propia.....	30
Figura 3.3 Vista isométrica frontal y vista trasera del prototipo Elaboración propia.....	31
Figura 3.4 Vista frontal en 3D del prototipo final de módulo de la red mínima de monitoreo de la contaminación del aire interior. Elaboración propia	32
Figura 3.5 Vista trasera en 3D del prototipo final de módulo de la red mínima de monitoreo de la contaminación del aire interior. Elaboración propia.....	32
Figura 3.6 Representación gráfica de sensibilidad utilizada para convertir la señal de salida del sensor en una proporción de partes por millón de CO.....	34
Figura 3.7 Representación gráfica de sensibilidad utilizada para convertir la señal de salida del sensor en una proporción de partes por millón de CO_2	35
Figura 3.8 Ubicación del módulo dentro del ambiente de implementación. Elaboración propia.....	36
Figura 3.9 Prototipo en modo de inicialización.....	37
Figura 3.10 Prototipo detectando aire limpio.....	37
Figura 3.11 Prototipo detectando calidad del aire regular.....	38
Figura 3.12 Prototipo detectando calidad del aire mala.....	38
Figura 3.13 Prototipo detectando calidad del aire muy mala.....	39
Figura 3.14 Prototipo detectando calidad del aire extremadamente mala.....	39
Figura 3.15 Grafico de caja de los datos de $PM_{2.5}$	45
Figura 3.16 Grafico de caja de los datos de PM_{10}	45
Figura 3.17 Grafico de caja de los datos de CO.....	46
Figura 3.18 Grafico de caja de los datos de CO_2	46
Figura 3.19 Grafico de caja de los datos de temperatura.....	47
Figura 3.20 Grafico de caja de los datos de humedad.....	47
Figura 3.21 Gráfico de caja para los datos de $PM_{2.5}$ y PM_{10} tras la exclusión de valores atípicos.....	49
Figura 3.22 Grafico de caja de los datos de CO sin valores atípicos	50
Figura 3.23 Datos históricos mensuales de $PM_{2.5}$	52
Figura 3.24 Descomposición estacional de $PM_{2.5}$	52
Figura 3.25 Datos históricos mensuales de PM_{10}	53
Figura 3.26 Descomposición estacional de PM_{10}	53
Figura 3.27 Datos históricos de CO	54
Figura 3.28 Descomposición estacional de CO.....	54
Figura 3.29 Datos históricos mensuales de CO_2	55
Figura 3.30 Descomposición estacional de CO_2	55

Figura 3.31 Datos históricos mensuales de Temperatura.....	56
Figura 3.32 Descomposición estacional de temperatura.....	56
Figura 3.33 Datos históricos mensuales de humedad.....	57
Figura 3.34 Descomposición estacional de humedad.....	57
Figura 3.35 Estimación de densidad de $PM_{2.5}$	58
Figura 3.36 Diagrama de densidad de PM_{10}	58
Figura 3.37 Diagrama de densidad de CO	59
Figura 3.38 Diagrama de densidad de CO_2	59
Figura 3.39 Diagrama de densidad de temperatura.....	60
Figura 3.40 Diagrama de densidad de humedad.....	60
Figura 3.41 Diagrama de dispersión de PM_{10} con CO	61
Figura 3.42 Diagrama de dispersión de PM_{10} con $PM_{2.5}$	61
Figura 3.43 Diagrama de dispersión de PM_{10} con CO_2	62
Figura 3.44 Diagrama de dispersión de PM_{10} con humedad.....	62
Figura 3.45 Diagrama de dispersión de PM_{10} con temperatura.....	63
Figura 3.46 Matriz de correlaciones entre las distintas magnitudes.....	63
Figura 3.47 Predicción mensual de $PM_{2.5}$	66
Figura 3.48 Predicción mensual de PM_{10}	66
Figura 3.49 Predicción mensual de CO	67
Figura 3.50 Predicción mensual de CO_2	67
Figura 3.51 Árbol de decisión potenciado por gradientes de PM_{10} y $PM_{2.5}$ parte 1.....	68
Figura 3.52 Árbol de decisión potenciado por gradientes de PM_{10} y $PM_{2.5}$ parte 2.....	69
Figura 3.53 Árbol de decisión potenciado por gradientes de CO parte 1.....	71
Figura 3.54 Árbol de decisión potenciado por gradientes de CO parte 2.....	72
Figura 3.55 Árbol de decisión potenciado por gradientes de CO_2 parte 1.....	74
Figura 3.56 Árbol de decisión potenciado por gradientes de CO_2 parte 2.....	75
Figura 4.1 Comparación entre datos reales de $PM_{2.5}$ y el método ARIMA.....	81
Figura 4.2 Comparación entre datos reales de PM_{10} y el método ARIMA.....	81
Figura 4.3 Comparación entre datos reales de CO y el método ARIMA.....	82
Figura 4.4 Comparación entre datos reales de CO_2 y el método ARIMA.....	82

Índice de tablas

Tabla 1.1 Monitoreo de partículas de contaminantes suspendidos en el aire interior.	12
Tabla 1.2 Monitoreo de condiciones meteorológicas.	12
Tabla 3.1 Sensores usados para el módulo de la red de monitoreo del aire interior.	28
Tabla 3.2 Modelo de datos obtenido.	40
Tabla 3.3 Modelo de datos definitivo.	41
Tabla 3.4 Porcentaje de datos vacíos.	42
Tabla 3.5 Porcentaje de datos vacíos en el marco de datos base del análisis.	43
Tabla 3.6 6 Datos estandarizados.	44
Tabla 3.7 Resultados Dickey-fuller sobre los datos del año 2022 de los contaminantes suspendidos en el aire interior, así como sus condiciones climáticas.	64
Tabla 3.8 modelos ARIMA definitivos que se escogieron de acuerdo con los criterios de ajuste.	65
Tabla 3.9 correlación de variables de acuerdo con árboles de decisión.	77
Tabla 4.1 Resultados de las predicciones futuras de los contaminantes del aire interior. .	80
Tabla 4.2 Resultados de las predicciones futuras de los contaminantes del aire interior. .	83

CAPÍTULO I. ANTECEDENTES

1.1 Introducción

La contaminación del aire interior es un desafío ambiental de creciente preocupación en todo el mundo. A medida que la sociedad moderna se adentra en un estilo de vida cada vez más orientado hacia espacios cerrados, como hogares, oficinas, escuelas y centros comerciales, la calidad del aire que respiramos en estos entornos adquiere una importancia crítica para nuestra salud y bienestar. Aunque a menudo tendemos a asociar la contaminación del aire con problemas en el exterior, como emisiones de vehículos o industrias, la realidad es que la calidad del aire en el interior de los edificios también puede estar plagada de una variedad de contaminantes que afectan directamente nuestra salud.

En contraste con la contaminación del aire exterior, que suele estar relacionada con fuentes identificables, la contaminación del aire interior puede tener múltiples orígenes, algunos de los cuales pueden pasar desapercibidos a simple vista. Esta se deriva de la interacción de una amplia gama de factores, como la elección de materiales de construcción, la presencia de productos químicos en muebles y productos de limpieza, así como la utilización de sistemas de climatización y calefacción. Además, actividades cotidianas aparentemente inofensivas, como cocinar, fumar, usar productos de cuidado personal y quemar velas, pueden contribuir significativamente a la acumulación de contaminantes en el aire interior.

Los contaminantes del aire interior abarcan desde partículas suspendidas en el aire hasta compuestos químicos volátiles, y pueden tener un impacto profundo en la salud humana. La exposición crónica a estos contaminantes ha sido relacionada con una serie de efectos adversos para la salud, incluyendo la exacerbación de enfermedades respiratorias existentes, el desarrollo de problemas cardiovasculares y en algunos casos, el aumento del riesgo de padecer cáncer. Además, ciertos grupos demográficos, como niños pequeños, ancianos y personas con afecciones médicas preexistentes, son particularmente susceptibles a los efectos negativos de la contaminación del aire interior.

La resolución de este problema ambiental requiere una combinación de esfuerzos y medidas preventivas. Una de las estrategias clave es mejorar la ventilación en los espacios interiores para garantizar un flujo adecuado de aire fresco. La instalación de sistemas de ventilación eficientes y la promoción de la ventilación natural pueden contribuir significativamente a la reducción de los niveles de contaminantes. Además, la elección de materiales de construcción y muebles de bajo contenido en compuestos químicos volátiles, así como la adopción de productos de limpieza y productos de cuidado personal más seguros, también son pasos esenciales para mitigar la contaminación del aire interior.

La educación pública y la concienciación sobre los peligros de la contaminación del aire interior también desempeñan un papel fundamental en la promoción de cambios de comportamiento. Informar a las personas sobre las fuentes de contaminación y los riesgos asociados puede empoderar a las comunidades para tomar decisiones informadas sobre la calidad del aire en sus hogares y lugares de trabajo. Las campañas de sensibilización pueden motivar a la adopción de prácticas más saludables, como la ventilación regular de espacios interiores, la reducción del uso de productos químicos agresivos y la implementación de cambios en la rutina diaria para minimizar la generación de contaminantes.

La búsqueda de ambientes interiores saludables y libres de contaminantes se ha elevado a la vanguardia de la conciencia colectiva, impulsando una necesidad urgente de comprender y abordar la calidad del aire que respiramos en nuestros hogares, oficinas y otros lugares cerrados.

El presente documento, ayuda a comprender cómo la elección de materiales de construcción, las actividades humanas, los sistemas de ventilación y climatización, así como otros factores, pueden influir en la composición del aire que nos rodea en espacios cerrados. También analiza los riesgos para la salud asociados con una mala calidad del aire interior y cómo estas preocupaciones se han convertido en un motor de cambio en la forma en que diseñamos, construimos y habitamos nuestros edificios.

En esta tesis, se descubrirá cómo el análisis detallado del aire interior se ha convertido en una herramienta esencial para evaluar y comprender la presencia de contaminantes, así como para guiar decisiones informadas sobre la implementación de mejoras y soluciones. Se implementará un prototipo de un módulo de una red mínima que analice la calidad del aire interior con mediciones precisas y monitoreo continuo.

En última instancia, este documento tiene como objetivo arrojar resultados que ayuden a comprender la importancia crítica de la calidad del aire interior en nuestra vida diaria.

1.2 Estado del arte

En esta sección se presentan las investigaciones más destacadas, así como las tecnologías, enfoques y avances más relevantes con relación a los estudios de calidad del aire interior.

1.2.1 Estimación de la calidad del aire dentro de diferentes laboratorios químicos

El objetivo de este proyecto de investigación fue evaluar la calidad del aire en los laboratorios químicos de la Universidad Técnica de Ambato utilizando Redes Neuronales Artificiales (RNA). Se implementó un modelo de RNA mediante el algoritmo de Retro propagación, que organiza las capas intermedias entre la entrada y la salida de la red. Se llevaron a cabo mediciones en 25 laboratorios con altos niveles de dióxido de carbono. Se emplearon dispositivos como el MX6 iBRID para medir monóxido y dióxido de carbono de forma directa, el Extech HT30 para medir temperatura y humedad, y el contador de partículas portátil Aerocet 831 para medir material particulado. Los puntos de muestreo se seleccionaron siguiendo la norma UNE 171330-2:2014. Se configuró una red en cascada con 2 capas ocultas, utilizando un método de entrenamiento basado en el descenso de gradiente, durante un total de 7 iteraciones. Esto condujo a un rendimiento del 80 por ciento. Se obtuvo una línea de tendencia lineal y un valor de $R=0.99557$, lo que indica que las 7 iteraciones de entrenamiento son óptimas para evitar fallos en la red neuronal (Gavilánez, 2021).

1.2.2 Prototipo de bajo costo para monitoreo del aire en ambientes cerrados

En este documento de investigación se presenta la contextualización, el diseño y la implementación de un prototipo económico para monitorear la calidad del aire en espacios interiores. Se llevaron a cabo mediciones de gases contaminantes en los laboratorios de telecomunicaciones de la Universidad Piloto de Colombia, donde se llevan a cabo diversos proyectos que implican el uso de productos que generan emisiones contaminantes, como soldaduras, pinturas y elementos eléctricos que se calientan y pueden generar partículas que quedan suspendidas en el aire. El

objetivo del proyecto fue realizar un análisis de la calidad del aire en los laboratorios de telecomunicaciones para evaluar la toxicidad del entorno (CRISTIAN, 2019).

1.2.3 Sistema de sensores que detecta sustancias peligrosas en el aire

Este sistema tiene la capacidad de analizar sustancias peligrosas presentes tanto en entornos laborales como domésticos. Se compone de tres subsistemas de detección independientes que miden tres parámetros diferentes y activan una alarma correspondiente en caso de detectar una situación de riesgo. Además, cuenta con comunicación inalámbrica a través de la cual los datos medidos son transmitidos a una estación de control. Estos subsistemas se han diseñado de forma conjunta, pero manteniendo su independencia entre sí, lo que permite adaptar el detector a diferentes parámetros de medición o utilizarlo en otros entornos de operación reemplazando los subsistemas correspondientes (Menéndez, 2015).

1.2.4 Uso de redes inalámbricas de sensores para monitorear gases contaminantes en espacios cerrados

El objetivo principal de este proyecto es desarrollar un sistema de monitoreo de gases contaminantes en entornos cerrados, con el propósito de facilitar la toma de decisiones por parte de los responsables del hogar. Para identificar la presencia de gases contaminantes y transmitir la información a través de la aplicación móvil Blynk, se propone una red de sensores basada en Arduino. Con la ayuda de esta aplicación, se pueden generar informes y alertas para ayudar en la implementación de medidas preventivas. De acuerdo con la gravedad de cada caso, los resultados obtenidos demuestran un alto porcentaje del correcto funcionamiento del prototipo, con respuestas en tiempo real y certeras (Revista Universidad y Sociedad, 8 de Abril del 2020).

1.2.5 Un sistema de monitoreo de seguridad de construcción en tiempo real para gases peligrosos que integra redes de sensores inalámbricos y tecnologías de modelado de información de edificios

Este documento presenta la integración de Building Information Modeling (BIM) y redes de sensores inalámbricos (WSN) en un sistema único que permite al sitio de construcción monitorear visualmente el estado de seguridad mediante una interfaz espacial de colores y eliminar automáticamente cualquier gas peligroso. Se colocaron numerosos nodos de sensores inalámbricos en un sitio de construcción subterráneo para recolectar datos sobre el nivel de gases peligrosos y las condiciones ambientales como temperatura y humedad. Cuando se detecta una condición anormal en una región específica, el modelo BIM generará una alerta y se activarán una alarma y un ventilador en el sitio para advertir y eliminar el peligro de forma automática. Este sistema propuesto tiene el potencial de mejorar significativamente la eficiencia en la gestión de la seguridad en la construcción y proporcionar información valiosa para las operaciones de rescate. Además, se presenta un estudio de caso que demuestra la aplicabilidad del sistema propuesto, y se resumen los beneficios prácticos, limitaciones, conclusiones y recomendaciones para futuras aplicaciones (Cheung, 2018).

1.3 Planteamiento del problema

Según los datos de la OPS (Organización panamericana de la salud, 2018), Las partículas suspendidas en el aire, como $PM_{2.5}$, PM_{10} , dióxido de carbono, monóxido de carbono, gas butano, plomo y dióxido de azufre, son los principales contribuyentes a la contaminación del aire en espacios cerrados, como casas, oficinas, fábricas, teatros y escuelas, entre otros.

Estas partículas tienen efectos en la salud muy graves y son capaces de penetrar ampliamente en las vías respiratorias, por lo que generan el agravamiento de enfermedades de ámbito respiratorio y cardiovascular.

Debido a que no se cuenta con la tecnología para monitorear y conocer los niveles de contaminación a los que están expuestas las personas en ambientes cerrados, se toman medidas de precaución con dispositivos que filtran el aire, pero estos pueden generar iones y ozono que en algunos casos les puede afectar a las personas, además dichos dispositivos solo limpian el aire del interior, pero no detectan el tipo de contaminantes a los que están expuestas las personas, siendo la exposición a los contaminantes una de las principales causas de muerte prematura en el mundo y un riesgo evitable a la salud.

1.4 Objetivos

1.4.1 Objetivo general

Realizar el prototipo de una red mínima de monitoreo de contaminantes en ambientes cerrados la cual, detecte las emisiones de partículas suspendidas en el aire producidas por contaminantes como: $PM_{2.5}$, PM_{10} y gases tóxicos.

1.4.2 Objetivos específicos

- Construir un prototipo de red de sensores de monitoreo de contaminantes simple, económica y efectiva que pueda detectar partículas de gases contaminantes en espacios cerrados.
- Utilizar el código ARDUINO, que permite la recopilación, copia de seguridad y visualización de datos, para probar la funcionalidad de las diversas partes del prototipo.
- Usar el prototipo en un ambiente cerrado para examinar cómo funciona para detectar niveles de gases tóxicos y contaminantes suspendidos en el aire.
- Predecir los niveles de contaminantes presentes en el aire interior utilizando algoritmos de aprendizaje automático.

1.5 Hipótesis

Es posible crear una red mínima para monitoreo de contaminantes que permita detectar la presencia de material particulado PM y gases tóxicos que se encuentren suspendidos en el aire dentro de ambientes cerrados, permitiendo la recolección de datos atmosféricos para la predicción del nivel de los contaminantes a los que se exponen diariamente las personas en cualquier ambiente cerrado.

1.6 Justificación

1.6.1 Justificación social

Al dar a conocer a la sociedad los niveles de contaminación a los que las personas se exponen diariamente en los lugares donde pasan la mayor parte de su tiempo como en escuelas, oficinas, fabricas, casas, restaurantes, entre otros, será de gran ayuda para que se puedan realizar evaluaciones de riesgos y así saber los efectos que tienen en la salud los contaminantes del aire que se encuentran el interior, para buscar soluciones a estos y tomar las medidas de adaptación o precaución necesarias para monitorear la calidad del aire interior que respiran.

1.6.2 Justificación tecnológica

El desarrollo de tecnología propia abre la puerta a la adquisición de un conocimiento en profundidad acerca de la composición de las mezclas de contaminantes que se originan en el aire interior. Estos contaminantes se presentan en concentraciones que varían, tanto en niveles altos como bajos, y su impacto se ve afectado por factores como la humedad, la temperatura y otros elementos del aire exterior que ingresan en los espacios cerrados, donde las personas pasan la mayor parte de su tiempo. Asimismo, se integra en este proceso la asimilación tecnológica, la cual define cómo las personas, las organizaciones y las sociedades adoptan y se adaptan a este tipo de tecnologías innovadoras.

1.6.3 Justificación económica

La percepción de las personas sobre la calidad del aire interior se puede mejorar creando una red mínima para monitorear contaminantes en aire interior a un costo menor que los altos costos de compra e instalación de equipos comerciales disponibles actualmente en el mercado.

1.7 Variables, indicadores e instrumentos

1.7.1 Variable dependiente

Los módulos que conformen la red de monitoreo de contaminantes serán dependientes del área en la que sean ubicados.

1.7.2 Variables independientes

En la Tabla 1.1, se muestran las variables independientes y sus indicadores; que se derivaron de la investigación realizada (ver Tabla 1.1).

Tabla 1.1 Monitoreo de partículas de contaminantes suspendidos en el aire interior.

Monitoreo de partículas de contaminantes suspendidos en el aire interior		
Variable	Indicador	Unidad de medida
$PM_{2.5}$	Respiración rápida y falta de oxígeno	$\mu\text{g}/\text{m}^3$
PM_{10}	Respiración rápida y falta de oxígeno	$\mu\text{g}/\text{m}^3$
Dióxido de carbono: CO_2	Respiración rápida y falta de oxígeno	$\mu\text{g}/\text{m}^3$
Monóxido de carbono: CO	Respiración rápida y falta de oxígeno	Ppm o mg/m^3

Tabla 1.2 Monitoreo de condiciones meteorológicas.

Otras variables para medir: humedad y temperatura		
Variable	Indicador	Unidad de medida
Temperatura	Temperatura del aire	$^{\circ}\text{C}$
Humedad	Vapor de agua	1 m^3

La Tabla 1.2 muestra las variables que servirán para monitorear las condiciones meteorológicas del lugar donde se ubique el módulo de la red (ver Tabla 1.2).

1.8 Limitaciones y Delimitaciones

1.8.1 Limitaciones

- Solo estará limitado a la recolección de los datos dentro de ambientes cerrados.
- Limitaciones otorgadas por el hardware y software utilizados para la elaboración del prototipo de la red de monitoreo de contaminantes en ambientes cerrados.

1.8.2 Delimitaciones

Se considerarán las siguientes delimitaciones en desarrollo del presente proyecto:

- Delimitación geográfica: solamente se aplica al área geográfica de Nogales, Sonora, México.
- Delimitación del conocimiento: la investigación se centra en el estudio e identificación de partículas de contaminantes del aire que se encuentran suspendidos dentro de ambientes cerrados.

1.9 Metodología

La metodología que se desglosa en las siguientes etapas, se desarrolló como parte de los conocimientos necesarios para administrar las actividades relacionadas con el desarrollo de una red mínima de monitoreo de la contaminación del aire interior.

1. Diseño y construcción:

En esta etapa se diseña la estructura donde se pondrá el circuito de la red de monitoreo de contaminantes, así mismo se determina el tipo de sensores que se utilizarán para el funcionamiento del prototipo.

2. Programación y pruebas:

En esta fase se identifican los modelos matemáticos para aplicar el ajuste necesario a los sensores de gases contaminantes, así como los algoritmos que ayudarán a la programación del hardware que compone el circuito de la red de monitoreo de contaminantes en ambientes cerrados.

Una vez lograda la fase de programación del prototipo de la red de monitoreo de contaminantes en ambientes cerrados, el paso siguiente consiste en evaluar su comportamiento y comparar su funcionamiento con los modelos matemáticos y los algoritmos implementados. Esta tarea tiene como finalidad exponer el módulo de la red a un ambiente real para ver su comportamiento al detectar las concentraciones de los gases contaminantes.

3. Origen, Evaluación y Extracción de los datos:

Este paso explica las características iniciales de los datos que se utilizaron y los pasos que se tomaron para finalmente combinarlos en un solo formato apropiado para la investigación y el análisis exploratorio.

4. Análisis exploratorio de los datos:

Para comprender los datos e identificar las posibles causas raíz del problema que este estudio intenta resolver, en esta fase se utilizan métodos estadísticos de resumen y herramientas gráficas

CAPÍTULO II. MARCO TEÓRICO

2.1 Terminología utilizada para el estudio de la calidad del aire interior

Los términos que se describen a continuación establecen un marco conceptual esencial para el desarrollo de la investigación propuesta.

2.1.1 Aire

El aire se define como el resultado de la mezcla de gases que componen la atmósfera terrestre, la cual es afectada por la gravedad en la superficie terrestre, y así mismo, el aire es el principal y más importante elemento para la vida en la Tierra (Ucha, 2019).

Su composición es el resultado de una combinación de diversas variables como nitrógeno (78%), oxígeno (21%), diversos vapores de agua (del 0% al 7%), dióxido de carbono, hidrógeno y gases nobles (1%).

2.1.2 Calidad del aire interior

La calidad del aire interior es el término utilizado para describir la calidad del aire dentro de los edificios y otras estructuras, especialmente en lo que se refiere al bienestar y la comodidad de los ocupantes del edificio. Cualquier sustancia o agente energético que pueda resultar en condiciones desfavorables, incluyendo gases como monóxido de carbono, radón y compuestos orgánicos volátiles, partículas y microorganismos contaminantes como moho y bacterias, pueden afectarlo (sostenibilidadedp, 2016).

2.1.3 Material particulado

El material particulado es una mezcla de partículas orgánicas e inorgánicas líquidas y sólidas suspendidas en el aire. Su composición es diversa, sus principales componentes son sulfato, nitrato, amoníaco, cloruro de sodio, carbón, polvo de mina, cenizas metálicas y agua.

Estas partículas también provocan reacciones químicas en el aire y se clasifican por su tamaño, en el área de calidad del aire hablamos de las partículas PM_{10} más

grandes con un diámetro aerodinámico teórico de 10 μm (una micra de un metro = una millonésima de metro) y partículas finas. Llamadas $PM_{2.5}$ que tienen 2.5 μm de diámetro. (Instituto para la salud geoambiental, 2020).

2.1.4 Gases tóxicos

Estos son gases que interactúan en los organismos vivos y pueden causar la muerte a una determinada concentración. Son causados por diversas fuentes, como: emisiones de vehículos, compuestos químicos de fábricas, polvo, polen y esporas de moho, que pueden suspenderse como partículas, y su inhalación aumenta la posibilidad de problemas de salud (Martinez, 26 de Agosto del 2014).

2.2 Técnicas utilizadas para el estudio de la contaminación del aire interior

Para lograr una comprensión más profunda de los patrones y las interacciones presentes en los datos, se recurre a diversas técnicas que ayuden en su estudio. Estas técnicas desempeñan un papel fundamental en la extracción de información de alta relevancia y en la guía de decisiones basadas en datos sólidos. Las técnicas descritas a continuación, son las más usuales utilizadas, en este tipo de estudios.

2.2.1 Análisis multivariado

El análisis multivariado se refiere a diferentes métodos para estudiar y probar los efectos simultáneos de varias variables. Los métodos estadísticos multivariados se utilizan para analizar el comportamiento general de varias variables aleatorias. Puede obtener información relevante al combinar métodos estadísticos que se enfocan en observar y manipular diferentes variables estadísticas simultáneamente. Existen dos amplias categorías de métodos de análisis estadístico multivariado, por un lado, los llamados métodos descriptivos y, por otro lado, los métodos explicativos (Crehana, 2021).

2.2.2 Diagrama de densidad

Los gráficos de densidad se pueden considerar como una extensión de los histogramas. A diferencia de los histogramas, los gráficos de densidad suavizan la distribución de valores y reducen el ruido. Los gráficos de densidad son los más

adecuados para determinar la forma de una distribución, el concepto de la forma de una distribución de probabilidad surge al tratar de encontrar una distribución adecuada para modelar las propiedades estadísticas de una población dada una muestra de esa población, porque no se ve afectada por el número de contenedores (Synergy codes, 2023).

2.2.3 Aprendizaje automático

El aprendizaje automático es un campo de la ciencia, específicamente una subdisciplina de la inteligencia artificial, que permite que los algoritmos detecten patrones repetitivos en conjuntos de datos. Cualquier cosa que se pueda almacenar digitalmente se puede usar como datos de aprendizaje automático. Al detectar patrones en estos datos, los algoritmos aprenden y mejoran su desempeño para una tarea en particular. Hay cuatro pasos principales para desarrollar un modelo de aprendizaje automático. **El primer paso** es tener y preparar un conjunto de datos de entrenamiento. Estos datos se utilizarán para alimentar un modelo de aprendizaje automático para descubrir cómo resolver el problema para el que fue diseñado. Los datos se pueden etiquetar para decirle al modelo qué características identificar. También pueden estar sin etiquetar, por lo que el modelo detectará y extraerá características duplicadas por sí mismo. En cualquier caso, los datos deben prepararse, organizarse y limpiarse cuidadosamente. Por otro lado, entrenar un modelo de aprendizaje automático puede estar sesgado. Los resultados de tus predicciones futuras tendrán un impacto directo, la selección de un algoritmo para ejecutar en el conjunto de datos de entrenamiento es **el segundo paso** y depende del tipo de problema que se está resolviendo, así como del tipo y la cantidad de datos de entrenamiento que se utilizarán. **El tercer paso** es entrenar el algoritmo. Es un proceso iterativo. Las variables se ejecutan a través del algoritmo y los resultados se comparan con lo que debería producir. Los pesos y las compensaciones se pueden ajustar para mejorar la precisión de los resultados. Luego, las variables se vuelven a ejecutar hasta que el algoritmo proporcione el resultado correcto en la mayoría de los casos. Utilizar el modelo y mejorarlo **es el último y cuarto paso**. (Datascientest, 2022).

2.2.4 Estacionalidad

La estacionalidad es una característica de las series temporales en las que los datos experimentan cambios frecuentes y predecibles, repitiéndose cada año calendario. Cualquier fluctuación o patrón previsto que se repita en el transcurso de un año se denomina estacional. El efecto estacional difiere del efecto cíclico en que los ciclos estacionales ocurren dentro de un año calendario, mientras que los efectos cíclicos, como el aumento de las ventas debido al bajo desempleo, pueden incluir períodos más cortos o largos que un año calendario (Kenton, 2020).

2.2.5 ARIMA

Los modelos integrados autorregresivos de media variable son un tipo de análisis de regresión que mide la fuerza de una variable dependiente en relación con otras variables. Al centrarse en las diferencias entre los valores de la serie en lugar de los valores reales, el modelo busca pronosticar valores futuros o cambios en los datos. Al delinear cada uno de sus elementos de la siguiente manera, se puede entender un modelo ARIMA.

Autorregresión (AR) : se refiere a un modelo en el que la variable que muestra el cambio se somete a una regresión utilizando su valor anterior o de retraso.

- **Integrado (I)**: este símbolo indica que las observaciones iniciales se han diferenciado para hacer que la serie temporal sea estacionaria, es decir el valor de los datos se ha reemplazado por la diferencia entre el valor de los datos y el valor anterior.
- **Media móvil (MA)** : combina la correlación de las observaciones con los residuos de un modelo de media móvil que se ha aplicado a las observaciones retrasadas.

2.2.6 Series temporales

Una serie de tiempo es una colección de observaciones realizadas en intervalos de tiempo que están espaciados sistemáticamente e incluye:

- **Tendencia:** El movimiento ascendente o descendente a largo plazo de una serie.
- **Estacionalidad:** Es la variación recurrente en una serie de tiempo durante un período específico. Estas variaciones crean un patrón que se repite con frecuencia de un período estacional al siguiente.
- **Ciclos:** Desviaciones prolongadas de la tendencia provocadas por elementos distintos a la estacionalidad. Los tiempos entre picos o valles sucesivos en un ciclo no siempre son iguales y los ciclos suelen abarcar un largo período de tiempo.
- **Movimiento irregular:** También conocido como ruido aleatorio o error en una serie temporal, es un movimiento que persiste incluso después de que se hayan tenido en cuenta los movimientos de tendencia, estacionales y cíclicos (Minitab, 2022).

2.2.7 Prueba de Dickey Fuller aumentada

Como prueba de significación estadística, la prueba de Dickey-Fuller aumentada arrojará resultados para pruebas de hipótesis que incluyen tanto la hipótesis nula como la alternativa. En consecuencia, tendremos un valor p a partir del cual determinar si la serie temporal es estacionaria o no.

La probabilidad conocida como valor p evalúa la solidez de los datos que respaldan la hipótesis nula. Las probabilidades más bajas proporcionan más evidencia en contra de la hipótesis nula.

Cuando el nivel de significación es 0.01, 0.05 o 0.10, la recomendación de análisis utiliza el valor crítico para probar la hipótesis nula porque el valor p contiene más aproximación. Tanto el valor crítico como el valor p generalmente conducen a la misma conclusión. La suposición de que los datos no son estacionarios implica que la diferenciación es un primer paso razonable para intentar que los datos sean estacionarios (Minitab, 2022).

2.3 Términos y herramientas de software utilizadas

En el análisis de datos, se utilizan una variedad de términos y herramientas de software para recopilar, limpiar, visualizar y analizar la información de manera efectiva. A continuación, se presentan algunos de los términos y herramientas más utilizados en esta investigación.

2.3.1 Datos

Los datos son números, letras o símbolos que describen cosas, circunstancias o eventos. Son la información fundamental sobre una persona, cosa o transacción que es relevante para una variedad de propósitos, incluida la toma de decisiones. Los datos se representan como pulsos electrónicos o pulsaciones desde la perspectiva de la computación mediante una red de circuitos conocida como señal digital. (Master magazine, 12 de Febrero del 2018).

2.3.2 Base de datos

Una base de datos es una colección de datos que se han configurado para ser fácilmente administrados, actualizados y accesibles.

En informática, las bases de datos se clasifican ocasionalmente en función de su enfoque organizativo. El enfoque que se usa con más frecuencia es la base de datos relacional, una base de datos tabular donde los datos se definen para que se puedan reorganizar y acceder a ellos de varias maneras. Una base de datos distribuida es aquella que se puede distribuir en varias ubicaciones en una red o replicarse allí. Una base de datos que soporta la programación orientada a objetos es aquella que es consistente con la información definida en las clases y subclases de objetos. (Rouse, 2018).

2.3.3 Excel Datastreamer

Un complemento COM para Microsoft Excel llamado Microsoft Data Streamer permite a los usuarios transmitir datos de baja latencia desde dispositivos y aplicaciones conectados a hojas de cálculo (Microsoft, 2022).

2.3.4 Minería de datos

La minería de datos es el proceso de obtener información útil de grandes conjuntos de datos. Puede identificar patrones y tendencias en los datos a través del análisis matemático porque, por lo general, las relaciones entre las variables son demasiado complejas o la cantidad de datos es demasiado grande para que la exploración de datos tradicional detecte estos patrones (Microsoft, 2019).

2.3.5 Python

Python es un poderoso lenguaje de programación de propósito general que es directo y fácil de aprender. Es un lenguaje de alto nivel que simplifica el trabajo con estructuras de datos numéricos y de texto. Python se usa con frecuencia en el análisis de datos y la extracción de información valiosa para las empresas. Junto con su simplicidad, que es una gran ventaja, Python tiene bibliotecas de procesamiento de datos como "Pydoop" que son muy beneficiosas para los expertos porque les permiten escribir código MapReduce en Python y procesar datos en un clúster.

La gestión y el análisis de datos se simplifican aún más con bibliotecas adicionales como dask y pyspark. La velocidad y la facilidad de escalado de Python lo hacen útil para crear información en configuraciones en tiempo real y convertir esa información a los lenguajes utilizados en big data. Además, existen otras bibliotecas como keras y tensorflow que son ricas en información sobre funcionalidades de aprendizaje automático (Soloaga, 2018).

2.4 Componentes electrónicos

La elección de los dispositivos electrónicos que se presentan a continuación responde a criterios de accesibilidad, sencillez de implementación, coste económico y una sólida capacidad para identificar con precisión los contaminantes presentes en el aire en entornos interiores.

2.4.1 Sensores

El prototipo de la red de contaminantes del aire interior debe poseer los siguientes sensores que le permitirán analizar las condiciones del aire interior, así como reconocer los objetos que se encuentren a su paso.

- **Sensor Mq-7 de monóxido de carbono(CO)**



Figura 2.1 Sensor MQ-7

En el rango de 20 a 2000 ppm, el sensor MQ-7 (ver Figura 2.1), puede detectar concentraciones de CO (Arduino, 2018).

- **Sensor de calidad del aire MQ-135**



Figura 2.2 Sensor de gases tóxicos

El MQ135 (ver Figura 2.2), es un sensor de calidad del aire que puede identificar gases nocivos como benceno, alcohol, dióxido de nitrógeno, amoníaco y monóxido de carbono. La principal aplicación del sensor es para la detección de gases peligrosos en la industria porque puede identificar concentraciones de gas. ya que detecta concentraciones de gas entre 10 y 1000 ppm. Puede activar rápidamente cualquier dispositivo porque responde lo suficientemente rápido (Arduino, 2018).

- **Sensor de partículas (Nova PM Sensor SDS011)**



Figura 2.3 Sensor de partículas (Nova PM Sensor SDS011)

El sensor de partículas SDS011 (ver Figura 2.3), puede determinar la concentración de partículas en el aire. Para determinar la concentración de partículas en su entorno, el sensor permite sus propias mediciones independientes. El número de partículas de 10 μm y 2,5 μm , que se definen como pautas de medición, se mide por cada metro cúbico de aire. (RadioShuttle, 2020).

- **Sensor de temperatura DHT11**



Figura 2.4 Sensor de humedad y temperatura.

El DHT11 es un sensor digital de temperatura y humedad relativa económico y fácil de usar (ver Figura 2.4), mide la humedad en el aire utilizando un termistor y un sensor de humedad capacitivo, y emite la información como una señal digital en el pin de datos. Tanto a nivel de software como de hardware, es muy fácil de usar (Arduino, 2018).

2.4.2 Arduino

En la plataforma de desarrollo Arduino, que se basa en una pieza de hardware libre, se incluyen un microcontrolador reprogramable y varios pines hembra. Estos hacen posible conectar el microcontrolador a los diversos sensores y actuadores en una variedad de formas muy fácilmente.

Con el fin de interactuar tanto con el hardware como con el software, Arduino se puede utilizar para crear elementos independientes, conectarse a otros dispositivos o comunicarse con otros programas. Se utiliza para transformar datos de una fuente, como un teclado, y convertir los datos en algo que entienda, como una computadora (Arduino, 2018). Como ejemplo, se puede usar un motor para subir o bajar una persiana en función de la cantidad de luz detectada por un sensor conectado al Arduino. (Arduino, 2018).

CAPÍTULO III. DESARROLLO

3.1 Fase de diseño y construcción

3.1.1 Arquitectura de la solución planteada

La Figura 3.1 representa el diseño de la arquitectura del sistema que integra una red elemental para el seguimiento de la calidad del aire en espacios interiores. Esta red se desglosa en tres partes: La primera parte se ocupa de la estructura fundamental del sistema y se basa en el sistema de control de sensores. Dicho sistema lleva a cabo la programación y comunicación de los sensores a través del puerto serial, permitiendo la transmisión de datos al sistema de adquisición de información. La segunda parte, representada en un recuadro gris, engloba la arquitectura del sistema electrónico, que proporciona la energía, el soporte y la protección necesarios tanto para el sistema de control de sensores como para los demás elementos que componen el circuito. Por último, la tercera parte de la solución, resaltada en color amarillo, corresponde al desarrollo de la arquitectura de almacenamiento de datos, encargada de recibir y gestionar los datos procedentes del sistema de comunicación serial con el propósito de su visualización e interpretación (ver Figura 31).

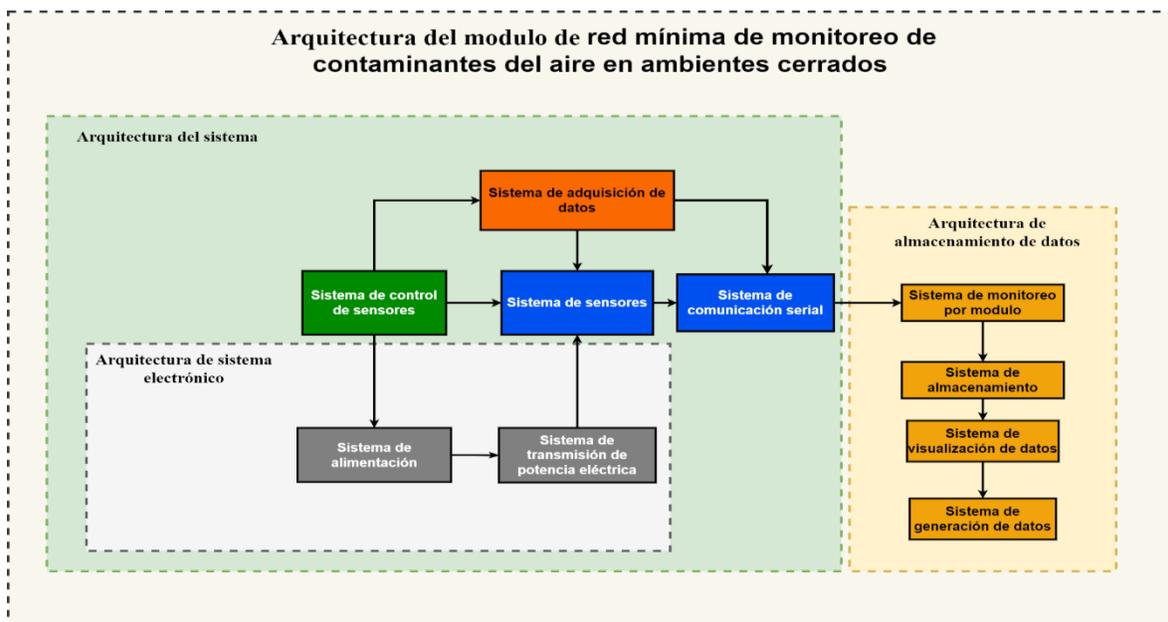


Figura 3.1 Arquitectura del sistema de monitoreo de calidad del aire en ambientes cerrados.

3.1.2 Arquitectura del sistema de sensores que incorpora el circuito

Para que el circuito de la red cumpla con su función principal de monitorear la calidad del aire interior debe ser capaz de recolectar y cuantificar información, que lo ayude a ejecutar sus acciones en el momento preciso. Para eso, se determinó que debía tener incorporados sensores que detectaran la presencia de partículas de polvo y gases contaminantes dentro del lugar donde se ubicará. La Tabla 3.1 muestra el tipo de sensores a utilizar y su función en específico.

Tabla 3.1 Sensores usados para el módulo de la red de monitoreo del aire interior.

Sensor	Función
MQ135 calidad del aire y MQ7 monóxido de carbono	Detecta gases peligrosos como Amoniaco, Dióxido de Nitrógeno, Alcohol, Benceno, Dióxido y Monóxido de carbono. En concentraciones de 10 y 1000 ppm
Nova PM SDS011	Mide la concentración de partículas de $PM_{2.5}$ y PM_{10} en el aire. Se mide por cada metro cúbico de aire, el número de partículas de 10 μm y 2,5 μm
DHT11	Hace mediciones de temperatura, humedad.

Los sensores mostrados en la Tabla 3.1 son semiconductores que cambian su conductividad y resistencia en función de la cantidad de $PM_{2.5}$, PM_{10} y gases contaminantes adsorbidos, poseen un circuito interno que funge como divisor de tensión. En general, el elemento semiconductor de estos se encuentra conectado en serie con una resistencia de carga (ver Tabla 3.1). De esta forma, cuando se aplica una diferencia de potencial entre el punto de salida y la tierra, es posible identificar el cambio de resistencia total en ellos, como una función de la cantidad de moléculas absorbidas sobre la superficie de los sensores.

3.1.3 Arquitectura de almacenamiento de datos

Para el almacenamiento de datos se utiliza un complemento de Excel llamado Data Streamer, el cual recopila y empaqueta en tiempo real los datos que envía el serial de Arduino. Los datos también se pueden guardar en un formato de valores separados por comas (CSV), lo que facilita su interpretación a la hora de hacer procedimientos estadísticos para redes neuronales.

3.1.4 Arquitectura de sistema electrónico

El sistema electrónico se compone de una fuente de poder con un rango de tensión de 9 V a 12 V regulando la tensión de entrada a 5 V, que es el voltaje de trabajo de los sensores y el Arduino. La arquitectura del sistema electrónico se encarga de abastecer de energía a la arquitectura del sistema la cual se compone de los sensores, los cuales envían información de las mediciones de partículas de CO_2 , CO, $PM_{2.5}$ y PM_{10} , así como medición de temperatura y humedad, a la arquitectura de almacenamiento de datos.

3.2 Definición de la estructura de los módulos de la red mínima de monitoreo de contaminantes de la calidad del aire interior

La estructura de los módulos detectores de emisiones de gases contaminantes en el aire interior es de gran importancia, ya que en la misma estarán montados todos los componentes, tales como: tarjetas electrónicas, baterías y sensores.

Para realizar un buen diseño de la estructura se analizan y enuncian todos los requerimientos básicos y específicos de la misma.

3.2.1 Requerimientos de la estructura

Los requerimientos básicos de la estructura son:

- El material deberá tener resistencia al peso de todos los componentes que estarán dentro de la estructura de los módulos.
- La estructura debe permitir el flujo de entrada y salida del aire para que sea más afectiva la detección de contaminantes, así como para evitar que se sobrecalienten los sensores.
- Debe de ser de un material ligero para facilitar su instalación y el traslado de un lugar a otro.
- El tamaño de la estructura está determinado por los elementos que se tienen que incorporar en su interior.
- La estructura tiene que cumplir una durabilidad suficiente para garantizar que se cumpla la función que los módulos tienen que desarrollar.

3.2.2 Diseño 2D de la estructura de los módulos de la red mínima de contaminantes de la calidad del aire interior

En base a los requerimientos de la estructura se determinó que debe ser una estructura sólida fabricada en material PETG, el cual es de excelente duración y resistencia, lo que hace que el prototipo tenga una forma adecuada que haga que el flujo de entra y salida del aire circule con facilidad por él, permitiendo que los sensores se activen y detecten la presencia de las partículas de contaminantes suspendidos en el aire interior, las Figuras 3.2 y Figura 3.3 muestran los planos isométrico y 3D de los módulos que conformarán la red de monitoreo de contaminantes suspendidos en el aire interior.

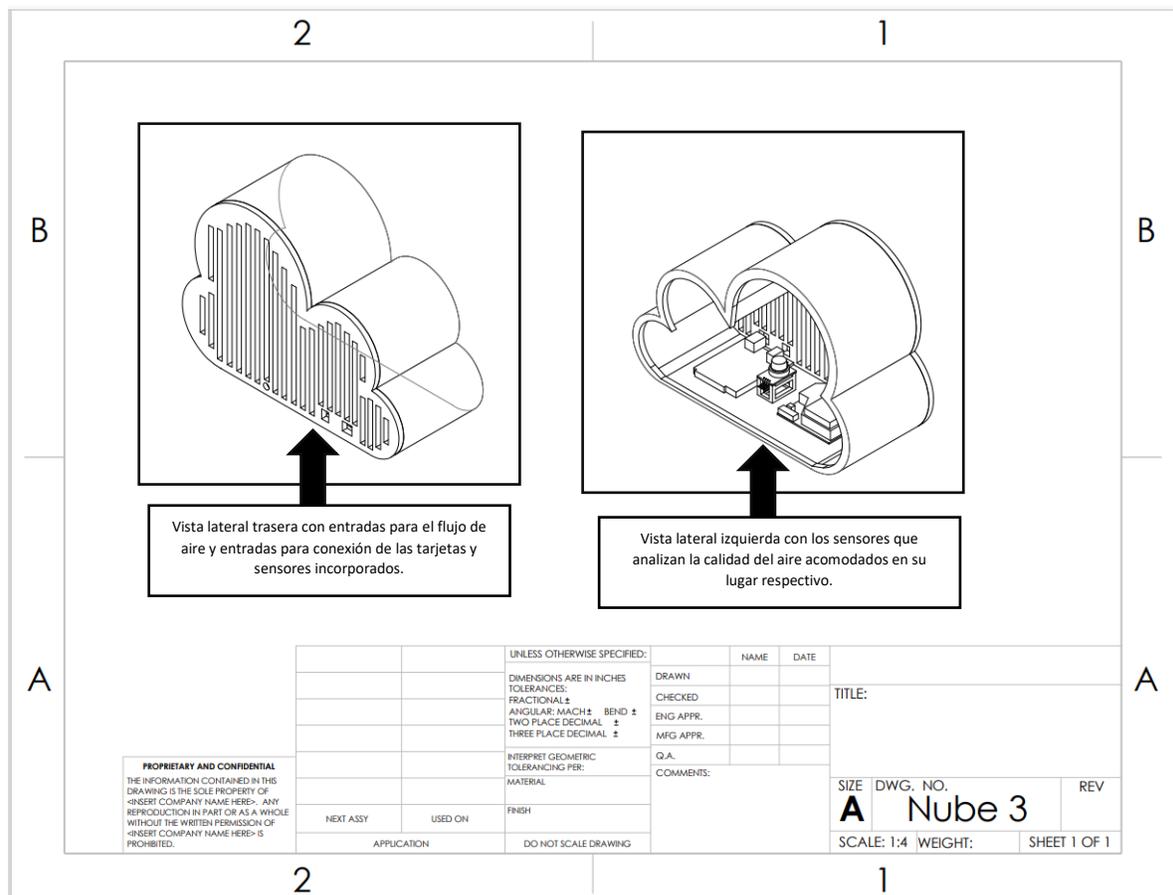


Figura 3.2 vista isométrica del diseño del prototipo de los módulos de la red de contaminantes del aire interior.

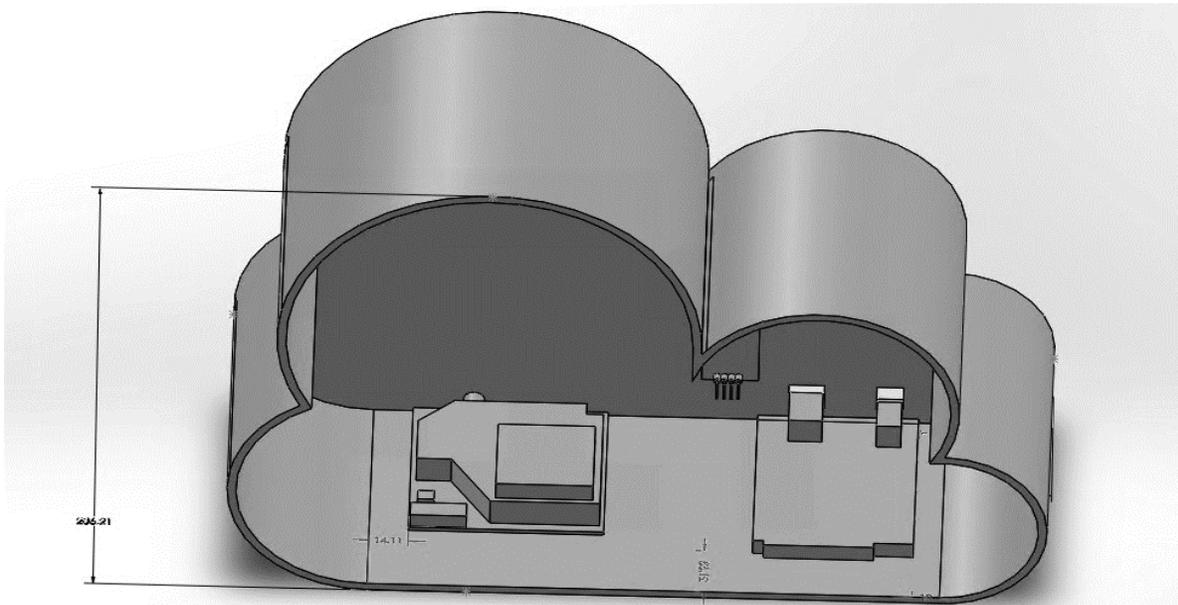


Figura 3.4 Vista frontal en 3D del prototipo final de módulo de la red mínima de monitoreo de la contaminación del aire interior.

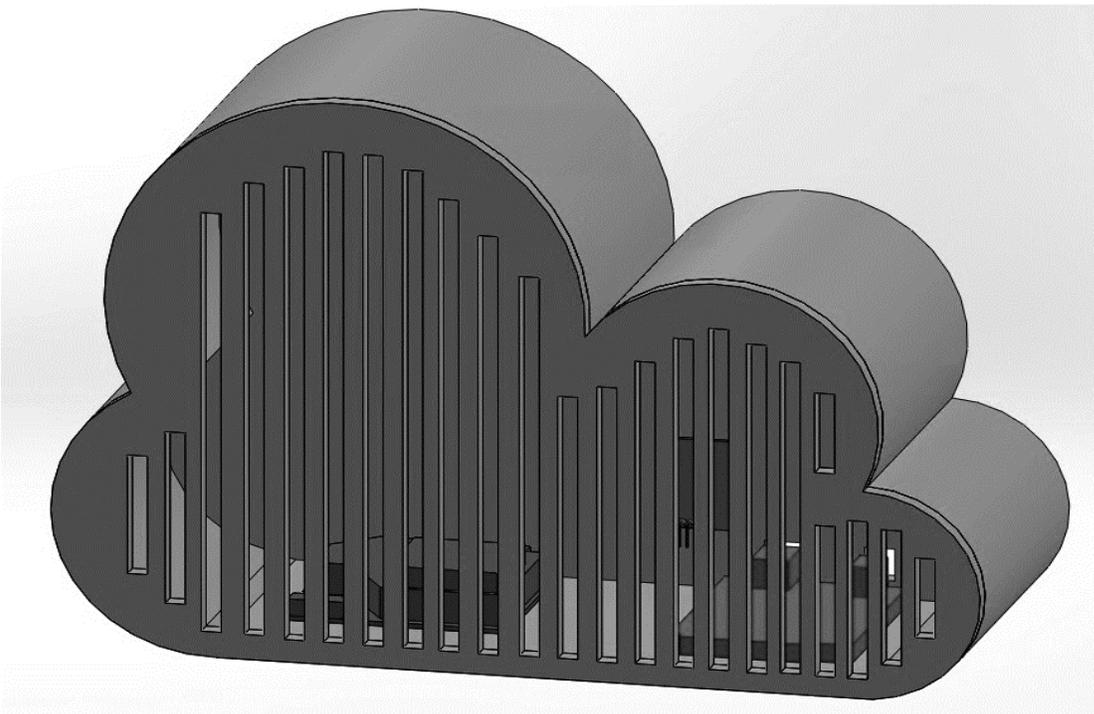


Figura 3.5 Vista trasera en 3D del prototipo final de módulo de la red mínima de monitoreo de la contaminación del aire interior.

3.3 Fase de programación y pruebas

3.3.1 Arranque y ajuste preciso de los sensores

Antes de que estos sensores puedan medir con precisión las concentraciones de los contaminantes, se debe realizar un procedimiento de activación que consta de dos fases:

El precalentamiento es la **primera** fase, que requiere dejar los sensores conectados a su fuente de alimentación de forma continua durante 24 horas, o incluso hasta 48 horas en el caso de los sensores de CO_2 y CO. Este proceso se utiliza para eliminar cualquier rastro de humedad o contaminación durante el proceso de fabricación y solo debe realizarse una vez.

La fase de ajuste preciso es la **segunda fase**, que implica averiguar qué tan lejos están los sensores de la realidad y usar esa información para ajustar las lecturas en la segunda fase. Se aplica el método de mínimos cuadrados para obtener medidas más precisas para las mediciones de los sensores para ayuda a identificar los valores de los parámetros que mejor se ajustan a los datos de prueba.

3.3.2 Calibración del sensor MQ-7

La calibración de un sensor MQ-7 implica ajustar el sensor para que proporcione lecturas precisas y confiables en función del gas que se está midiendo, en este caso, el monóxido de carbono (CO). La calibración se realiza en un ambiente donde se puedan controlar cuidadosamente las concentraciones de monóxido de carbono. Una opción es usar un área al aire libre con buena ventilación y sin riesgo para la salud.

Para lograr que las lecturas del sensor se ajusten a las condiciones ambientales se debe dejar que se estabilice en el entorno de calibración durante un tiempo específico aproximado de entre 24 y 48 horas.

Una forma de verificar si el sensor MQ-7 está recibiendo lecturas precisas es aplicando el método de la creación de la curva de calibración, la cual implica relacionar las lecturas del sensor con las concentraciones reales de CO.

La curva de ajuste del sensor MQ-7 se vería de la siguiente manera que se muestra en la Figura 3.6.

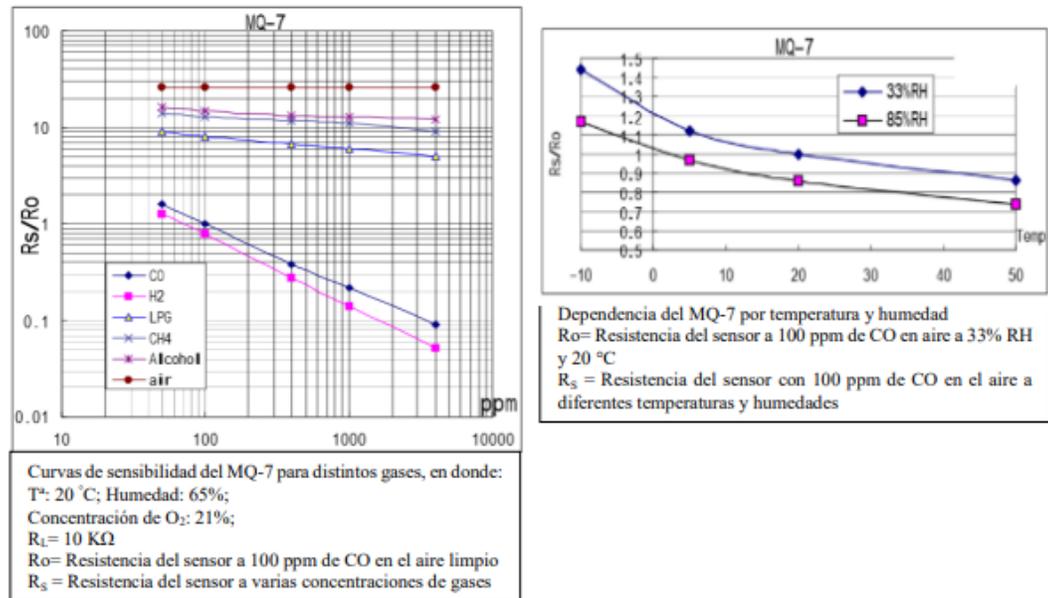


Figura 3.6 Representación gráfica de sensibilidad utilizada para convertir la señal de salida del sensor en una proporción de partes por millón de CO.

La figura 3.6 indica que la curva podría comenzar cerca del origen del gráfico (0,0) y luego subir gradualmente a medida que las concentraciones de CO aumentan (ver Figura 3.6).

En algunos casos, especialmente cuando se trata de sensores de bajo costo como el MQ-7, la curva podría no ser perfectamente lineal. Podría haber una región inicial donde la curva es más pronunciada, seguida de una región donde la relación entre las lecturas del sensor y las concentraciones de CO se suaviza.

Es importante recordar que la forma exacta de la curva de ajuste puede variar según diversos factores, como las condiciones ambientales, la calidad del sensor y otros factores específicos de configuración.

3.3.3 Calibración del sensor MQ135

La calibración del sensor MQ-135 implica ajustar el sensor para que proporcione lecturas precisas y confiables de CO_2 , se debe utilizar una fuente confiable y controlada del gas al que se va a ajustar el MQ-135. Esto podría ser un encendedor sin chispa que contenga mezclas conocidas del gas de interés.

Para que el sensor se ajuste a las condiciones ambientales se debe dejar que se estabilice en el ambiente de calibración durante un período de tiempo de 24 horas.

A este sensor también se le debe de aplicar el método de la curva de calibración relacionando las lecturas del sensor con las concentraciones reales del CO_2 como se muestra a continuación.

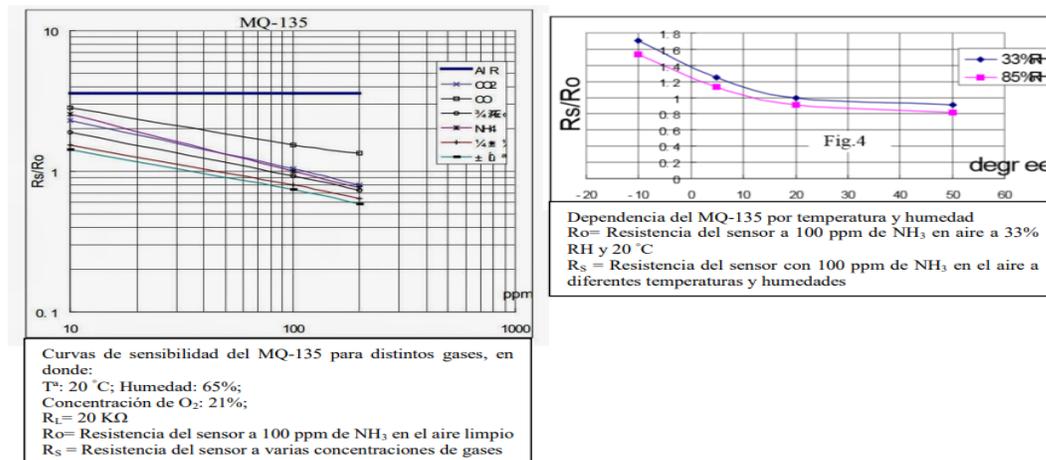


Figura 3.7 Representación gráfica de sensibilidad utilizada para convertir la señal de salida del sensor en una proporción de partes por millón de CO_2 .

De acuerdo con la Figura 3.7, la curva de calibración del sensor MQ-135 muestra la relación entre las lecturas del sensor y las concentraciones reales de un gas específico, como el dióxido de carbono en función de partes por millón. La curva podría tener una forma ascendente, donde a medida que aumenta la concentración de gas, las lecturas del sensor también aumentan. Por lo tanto, la pendiente y la forma de la curva dependerán de la sensibilidad y respuesta del sensor (ver Figura 3.7).

3.3.4 Calibración del sensor de temperatura y humedad DHT11

El sensor de temperatura y humedad puede requerir una calibración básica para asegurar mediciones precisas. Aunque no necesita una calibración exhaustiva como otros sensores, se debe de elegir un entorno que tenga condiciones estables de temperatura y humedad para dejar que se establezca durante un periodo de tiempo de 20 a 30 minutos para que se ajuste a las condiciones ambientales del lugar donde se ubicará.

3.3.5 Calibración del sensor SDS011 de partículas de PM_{10} y $PM_{2.5}$

La calibración del sensor SDS011, que es un sensor de partículas finas, es un proceso importante y para asegurarse de que las mediciones sean precisas y confiables, se debe seleccionar un lugar de calibración donde se puedan controlar las concentraciones de partículas de PM_{10} y $PM_{2.5}$ de manera precisa. Para ello se debe asegurar que el entorno esté bien ventilado, libre de otras fuentes de contaminación que puedan afectar las mediciones y dejarlo que se establezca durante un periodo de 30 minutos.

3.3.6 Implementación y funcionamiento del prototipo.

Para llevar a cabo la fase de pruebas, el módulo de monitorización se estableció en una ubicación fija dentro de la sala de un domicilio particular.

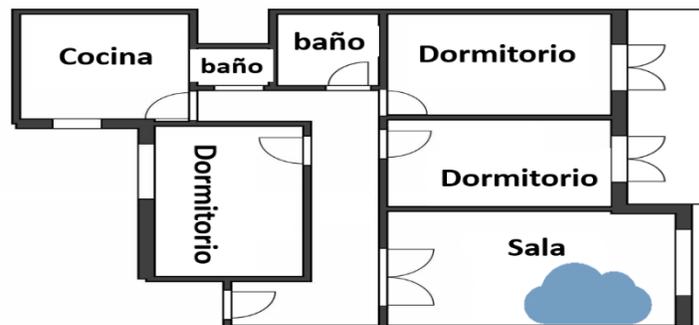


Figura 3.8 Ubicación del módulo dentro del ambiente de implementación.

Como se observa en la Figura 3.8, el módulo está en modo de inicialización recolectando datos de los contaminantes dentro del hogar (ver Figura 3.8).

3.3.7 Funcionamiento del prototipo

Las siguientes figuras muestran las diferentes fases por las que pasa el prototipo en su funcionamiento.



Figura 3.9 Prototipo en modo de inicialización.

La Figura 3.9 muestra el prototipo en color azul el cual significa que está en modo de inicialización y a la espera de empezar a detectar los contaminantes suspendidos en el aire interior (ver Figura 3.9).



Figura 3.10 Prototipo detectando aire limpio.

Como se observa en la Figura 3.10, cuando el prototipo se mantiene en color verde (ver Figura 3.10), significa que la calidad del aire interior es satisfactoria y existe poco o ningún riesgo para la salud por lo tanto se puede realizar cualquier actividad. Normal dentro del hogar.



Figura 3.11 Prototipo detectando calidad del aire regular.

Según la Figura 3.11 cuando el prototipo enciende de color amarillo (ver Figura 3.11), se detectan pocos niveles de contaminantes en el aire, por lo que la calidad del aire interior es aceptable, sin embargo, en el caso de algunos contaminantes, las personas que son inusualmente sensibles, pueden presentar síntomas moderados de alergias.



Figura 3.12 Prototipo detectando calidad del aire mala.

Al observar el comportamiento del prototipo en la Figura 3.12 se puede observar que se encendió de color naranja (ver Figura 3.12), lo cual significa que la calidad del aire interior es dañina para la salud de los grupos sensibles como niños, adultos mayores, personas que realizan actividad física intensa o con enfermedades respiratorias o cardiovasculares.



Figura 3.13 Prototipo detectando calidad del aire muy mala.

Según el comportamiento observado en la Figura 3.13, cuando el prototipo cambie a color rojo (ver Figura 3.13), significará que la calidad del aire interior es muy mala y dañina para la salud, por lo tanto, quienes pertenecen a los grupos sensibles pueden presentar efectos graves en su salud.



Figura 3.14 Prototipo detectando calidad del aire extremadamente mala.

Según la Figura 3.14, cuando el prototipo llegue al límite máximo permitido de los niveles de contaminantes suspendidos en el aire interior, cambiará a color morado el cual indica que la calidad del aire interior es extremadamente alta y muy dañina para la salud representando una condición de emergencia, ya que todas las personas que se encuentren en el lugar, se verían muy afectadas por este cambio drástico de la calidad del aire interior (ver Figura 3.14), cabe destacar que, en el interior, donde se puso a prueba el prototipo no llegó a niveles extremadamente altos, solo se representó el color morado de las luces para demostrar el funcionamiento del prototipo en un supuesto caso de emergencia.

3.4 Obtención, Interpretación y Extracción de Información

3.4.1 Procedencia de los datos

Los datos que se usan en este proyecto provienen de un módulo de la red de monitoreo del aire interior instalado en la sala de un hogar particular.

3.4.2 Obtención de los datos horarios de contaminación del aire interior

Para la obtención de los datos se dejaron los módulos de la red expuestos a un ambiente interior entre los meses de abril a julio del año 2022, en dicho lapso los módulos fueron detectando las concentraciones de gases tóxicos, así como partículas de polvo suspendidas en el aire interior. Para el proceso de obtención de datos se realizó el monitoreo por hora para cada variable medida.

3.4.3 Descripción de los datos obtenidos

La colección de datos que se obtuvo durante 4 meses cuenta con 9,834 datos, los cuales fueron extraídos del módulo de monitoreo de la calidad del aire interior ubicado dentro de un domicilio particular de la ciudad de Nogales, Sonora. El modelo de datos obtenido se muestra en la Tabla 3.2.

La Tabla 3.2 muestra las características que definen al conjunto de datos, las cuales son: fecha, $PM_{2.5}$, PM_{10} , CO_2 , CO, temperatura y humedad (ver Tabla 3.2).

Tabla 3.2 Modelo de datos obtenido.

Fecha	$PM_{2.5}$	PM_{10}	CO	CO_2	TEMP	HUM
1/1/2022 0:00	181	202	534	196	20.5	34.1
1/1/2022 1:00	53	82	536	136	20.5	34.3
1/1/2022 2:00	36	60	536	255	20.5	34.8
1/1/2022 3:00	143	188	536	244	20.5	35.1
1/1/2022 4:00	91	135	539	230	20.6	35.2
1/1/2022 5:00	32	55	543	232	20.6	35.3
1/1/2022 6:00	124	160	547	255	20.6	35.7
1/1/2022 7:00	89	114	546	267	20.6	36.1
1/1/2022 8:00	16	39	545	212	20.6	36
1/1/2022 9:00	32	56	545	225	20.6	36.2

3.4.4 Pre procesado de los datos

El primer paso para transformar los datos atmosféricos que se encuentran en crudo, en un formato más descriptivo es, separar de la fecha la hora, el día, mes y año en el que se tomó cada una de las mediciones de los contaminantes para que se obtenga un único marco de datos definitivo tal como lo muestra la Tabla 3.3.

Tabla 3.3 Modelo de datos definitivo.

Año	Mes	Día	Hora	$PM_{2.5}$	PM_{10}	CO	CO_2	TEMP	HUM
2022	1	1	0	181	202	534	196	20.5	34.1
2022	1	1	1	53	82	536	136	20.5	34.3
2022	1	1	2	36	60	536	255	20.5	34.8
2022	1	1	3	143	188	536	244	20.5	35.1
2022	1	1	4	91	135	539	230	20.6	35.2
2022	1	1	5	32	NA	543	232	20.6	35.3
2022	1	1	6	124	NA	547	255	20.6	35.7
2022	1	1	7	89	114	546	267	20.6	36.1
2022	1	1	8	16	NA	545	212	20.6	36
2022	1	1	9	32	56	545	225	20.6	36.2

La Tabla 3.3 presenta el conjunto de datos final que será empleado en las siguientes fases del proyecto (ver Tabla 3.3), quedando organizado de la siguiente manera:

- Año, mes, día y hora: momentos específicos en los que se obtuvieron los datos de los contaminantes del aire interior.
- $PM_{2.5}$: medida numérica que indica la cantidad de $PM_{2.5}$, partículas sólidas o líquidas presentes en la atmósfera con un diámetro inferior a 2.5 micrómetros, que se emitió a la hora y día indicados. Se mide en microgramos por metro cúbico ($\mu\text{g}/\text{m}^3$).
- PM_{10} : medida numérica que indica la cantidad de PM_{10} , partículas sólidas o líquidas presentes en la atmósfera con un diámetro comprendido entre 2.5 y 10 micrómetros, que se emitió a la hora y día indicados. Se mide en microgramos por metro cúbico ($\mu\text{g}/\text{m}^3$).
- CO: medida numérica que indica la cantidad de monóxido de carbono que se emitió a la hora y días indicados. Se mide en microgramos por metro cúbico ($\mu\text{g}/\text{m}^3$).

- CO_2 : medida numérica que indica la cantidad de dióxido de carbono que se emitió a la hora y día indicados. Se mide en microgramos por metro cúbico ($\mu g/m^3$).
- Temp: medida numérica que indica la temperatura del día en que se realizó el monitoreo.
- Hum: medida numérica que indica la humedad del lugar y día en que realizó el monitoreo.

3.4.5 Detección de los datos ausentes

Es una buena idea verificar si hay valores atípicos o datos faltantes antes de visualizar y analizar los datos y si existen, corregirlos adecuadamente. Como se mencionó en secciones anteriores, el análisis se comienza con el formato de datos de la Tabla 3.3, sin embargo, como se muestra en la Tabla 3.2, algunas mediciones no tienen datos disponibles por una variedad de razones, que incluyen:

- Errores en los instrumentos de medición.
- Problemas al procesar los datos.
- Problemas de pérdida de señal en los sensores.
- Error humano en el momento que se validaron esos datos.

Si faltan muchos datos, lo mejor es eliminar toda la magnitud; sin embargo, si faltan pocos datos, no es necesario llegar a ese punto. Por lo tanto, el primer paso es determinar cuántos datos faltan.

Tabla 3.4 Porcentaje de datos vacíos

Dato	Porcentaje de datos vacíos
Año	0
Mes	0
Día	0
Hora	0
$PM_{2.5}$	0.043234
PM_{10}	0.043234
CO	0.807033
CO_2	0
TEMP	0
HUM	0

Como se observa en la Tabla 3.4, al comprobar si los datos están incompletos, se puede notar que el número de datos vacíos es muy reducido, por lo cual se deben eliminar los datos faltantes de las columnas (ver Tabla 3.4). Esto se debe aplicar una sola línea de código por medio de la librería *pandas*.

Tabla 3.5 Porcentaje de datos vacíos en el marco de datos base del análisis.

Dato	Porcentaje de datos vacios
Año	0
Mes	0
Dia	0
Hora	0
$PM_{2.5}$	0
PM_{10}	0
CO	0
CO_2	0
TEMP	0
HUM	0

Al continuar con la limpieza y ordenamiento de datos, la Tabla 3.5 indica que no falta ninguno, lo cual tiene cierto sentido, porque los datos meteorológicos y atmosféricos ya fueron corregidos con la eliminación de los datos faltantes en la sección 3.4.5. A pesar de tener la completa seguridad de que a las mediciones no les pueden faltar datos, es una buena práctica comprobarlo para cerciorarse de que es así, para poder someterlos al proceso de estandarización (ver Tabla 3.5).

3.4.6 Estandarización de los datos

Los datos deben estar estandarizados para que todos estén dentro del mismo rango para permitir comparaciones entre las diversas magnitudes.

Si se quisiera comparar la evolución de la temperatura (con un valor máximo de alrededor de 40) y CO (con un valor máximo de alrededor de 1,616), el gráfico resultante sería completamente inútil. Este problema se resuelve estandarizando los datos, lo que centra el promedio de todas las magnitudes en 0 y establece la desviación estándar en 1. Esta estandarización se puede usar en una sola línea de código usando la librería de Python *scikit-learn*. Es crucial dejar en claro que solo se afectan las magnitudes numéricas; las mediciones cualitativas no se ven afectadas y deben convertirse en valores numéricos.

Cabe señalar que, a partir de ahora se utilizan todos los métodos y fórmulas para analizar la información de un solo módulo de la red de monitoreo de contaminación

del aire interior, que se encuentra alojado en el interior de una residencia privada en la ciudad de Nogales, Sonora.

Para mejorar los resultados de los algoritmos de aprendizaje automático, incluidos los algoritmos de agrupamiento y aprendizaje profundo, la estandarización se usa con frecuencia además de visualizar varias variables en la misma escala (ver Tabla 3.6).

Tabla 3.6 Datos estandarizados.

Año	Mes	Día	Hora	$PM_{2.5}$	PM_{10}	CO	CO_2	TEMP	HUM
2022	1	1	0	0.183027	0.235431	0.309313	0.279686	0.369295	0.503185
2022	1	1	1	0.052147	0.095571	0.311063	0.181551	0.369295	0.506369
2022	1	1	2	0.034765	0.06993	0.310811	0.376186	0.369295	0.514331
2022	1	1	3	0.144172	0.219114	0.310567	0.358194	0.369295	0.519108
2022	1	1	4	0.091002	0.157343	0.313825	0.335296	0.373444	0.520701
2022	1	1	5	0.030675	0.064103	0.317092	0.338567	0.373444	0.522293
2022	1	1	6	0.124744	0.18648	0.32035	0.376186	0.373444	0.528662
2022	1	1	7	0.088957	0.132867	0.319853	0.395813	0.373444	0.535032
2022	1	1	8	0.014315	0.045455	0.319348	0.305855	0.373444	0.533439
2022	1	1	9	0.030675	0.065268	0.318852	0.327118	0.373444	0.536624

La Tabla 3.6 muestra el modelo de datos estandarizados en un rango de entre 0 y 1 usando la librería de Python *scikit-learn*.

3.4.7 Detección y eliminación de datos atípicos

Para detectar estos valores atípicos en los datos, se utiliza una combinación de dos técnicas: los diagramas de caja y el rango intercuartílico.

El rango intercuartílico (IQR) se utiliza para medir la variabilidad, dividiendo un conjunto de datos en cuartiles. Los datos se ordenan en orden ascendente y se dividen en 4 partes iguales. Q1, Q2, Q3 llamados primer, segundo y tercer cuartiles, son los valores que separan las 4 partes iguales, donde:

- Q1 representa el percentil 25 de los datos.
- Q2 representa el percentil 50 de los datos.
- Q3 representa el percentil 75 de los datos.

IQR es el rango entre el primer y el tercer cuartil, es decir, $Q1$ y $Q3$: $IQR = Q3 - Q1$. Los puntos de datos que caen por debajo de $Q1 - 1.5 IQR$ o por encima de $Q3 + 1.5 IQR$ son valores atípicos.

Para visualizar los cuartiles de cada una de las magnitudes con sus valores atípicos se utilizan los diagramas de cajas antes mencionados.

- $PM_{2.5}$: A pesar de la uniformidad general de los datos, en la Figura 3.15. Se destacan algunos valores atípicos que sobresalen por encima del rango normal, afectando la distribución (ver Figura 3.15).

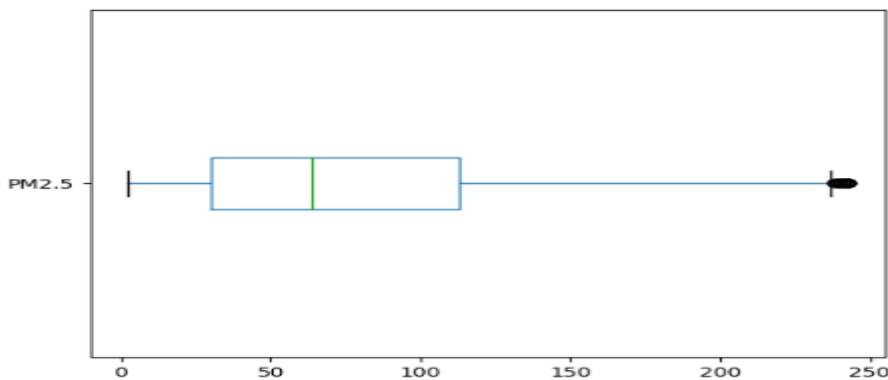


Figura 3.15 Grafico de caja de los datos de $PM_{2.5}$.

- PM_{10} : Aunque la mayoría de los datos exhibe una distribución uniforme, se observa en la Figura 3.16, un conjunto minoritario de valores con magnitudes significativamente superiores a los datos comunes (ver Figura 3.16), los cuales se encuentran fuera del rango habitual de concentraciones medidas de partículas de PM_{10} .

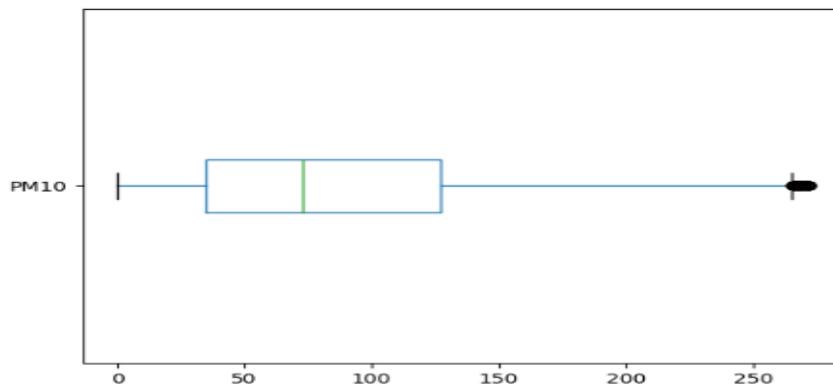


Figura 3.16 Grafico de caja de los datos de PM_{10}

- CO : Según la Figura 3.17, se identifica un grupo limitado de valores atípicos que se ubican fuera de los límites convencionales de los datos de monóxido de carbono (ver Figura 3.17).

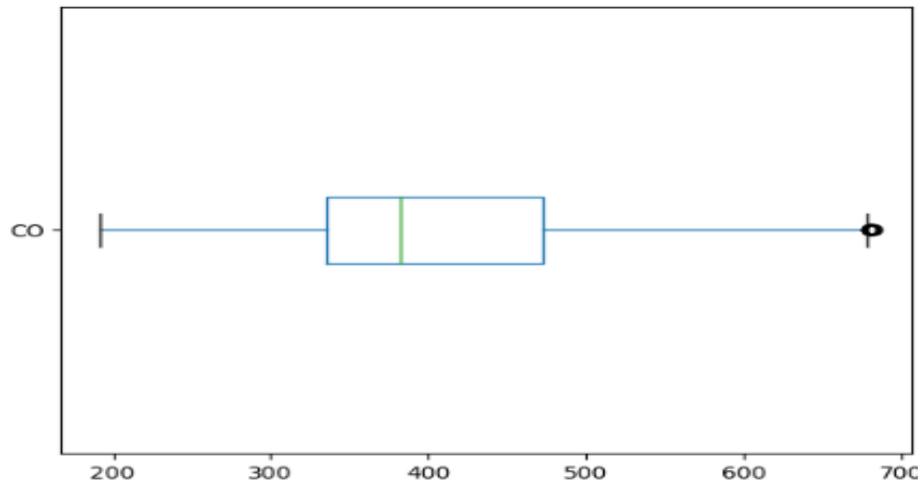


Figura 3.17 Grafico de caja de los datos de CO .

- CO_2 : no se encontró ningún valor atípico

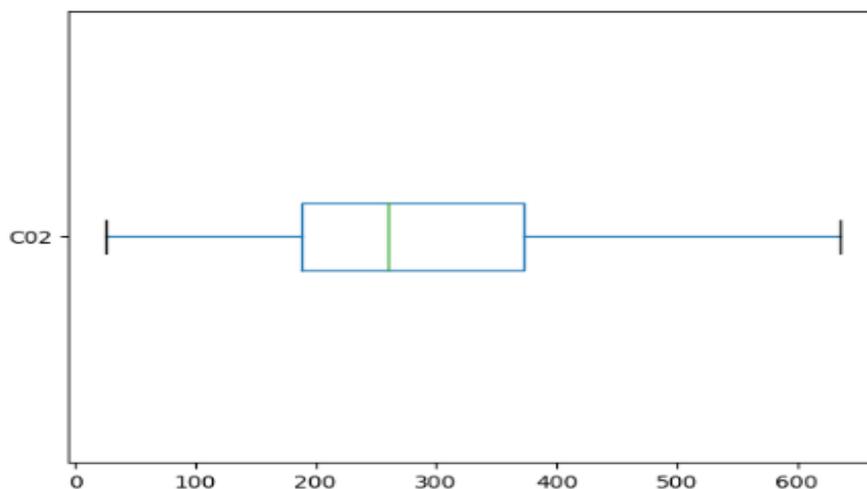


Figura 3.18 Grafico de caja de los datos de CO_2

De acuerdo con la Figura 3.18, no se identifica un grupo de datos atípicos fuera de límites convencionales de los datos de dióxido de carbono (ver Figura 3.18).

- Temperatura: no hay ningún valor atípico

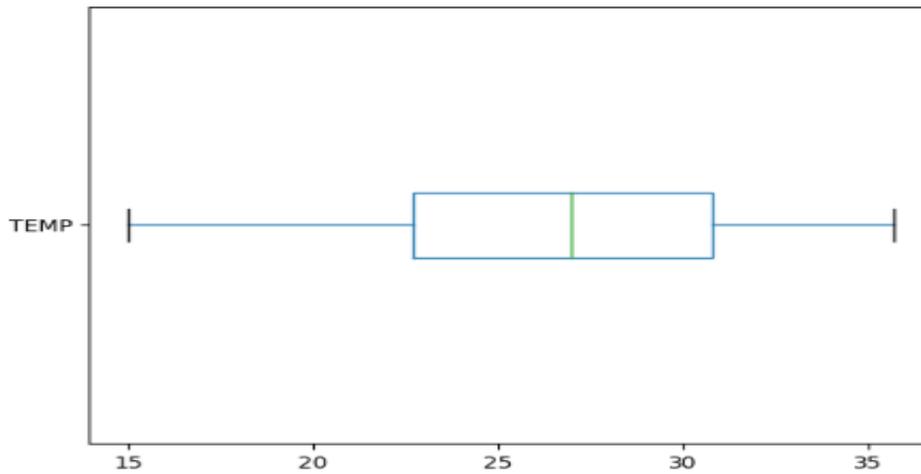


Figura 3.19 Grafico de caja de los datos de temperatura.

Como se observa en la Figura 3.19, no se identifica un grupo de datos atípicos fuera de límites convencionales de los datos temperatura (ver Figura 3.19).

- Humedad: no hay ningún valor atípico.

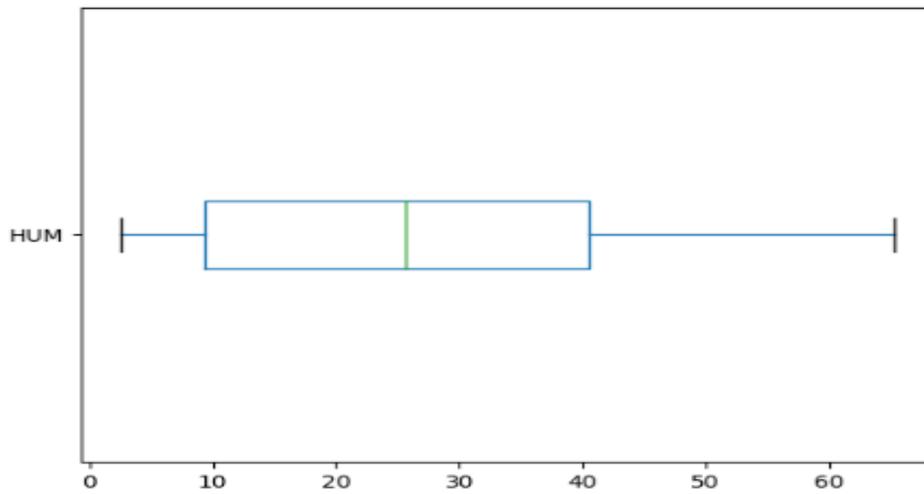


Figura 3.20 Grafico de caja de los datos de humedad.

De acuerdo con la Figura 3.20, no se identifica un grupo de datos atípicos fuera de límites convencionales de los datos de humedad (ver Figura 3.20).

3.4.8 Eliminación de datos atípicos

Tal como se puede apreciar en la sección 3.4.7, se presentan múltiples valores atípicos en diversas magnitudes de contaminantes, sin embargo, no se registran datos atípicos en las mediciones de las magnitudes meteorológicas.

Para la exclusión de los valores atípicos en las magnitudes de los contaminantes, se recomienda aplicar el mismo procedimiento que consiste en eliminar una entrada específica del conjunto de datos según su posición exacta en el mismo. Esto se debe a que, para la detección de valores atípicos, el resultado final se materializa como una lista que abarca todos aquellos elementos de datos que se ajustan a la definición de valores atípicos.

El fragmento de código `dataframe.drop(row_index, inplace=True)` se emplea para eliminar una fila específica del conjunto de datos utilizando los índices de fila que se desean suprimir. Al utilizar el parámetro `inplace=True`, se instruye a Python para que realice la modificación directamente en el conjunto de datos original. En cuanto a `row_index`, es posible asignar un solo valor, una lista de valores o incluso una matriz `NumPy`, pero es fundamental que se trate de una estructura unidimensional.

Para verificar nuevamente si la eliminación de todos los valores atípicos ha sido efectiva, se procede a emplear los diagramas de caja y a continuación se exponen los resultados obtenidos.

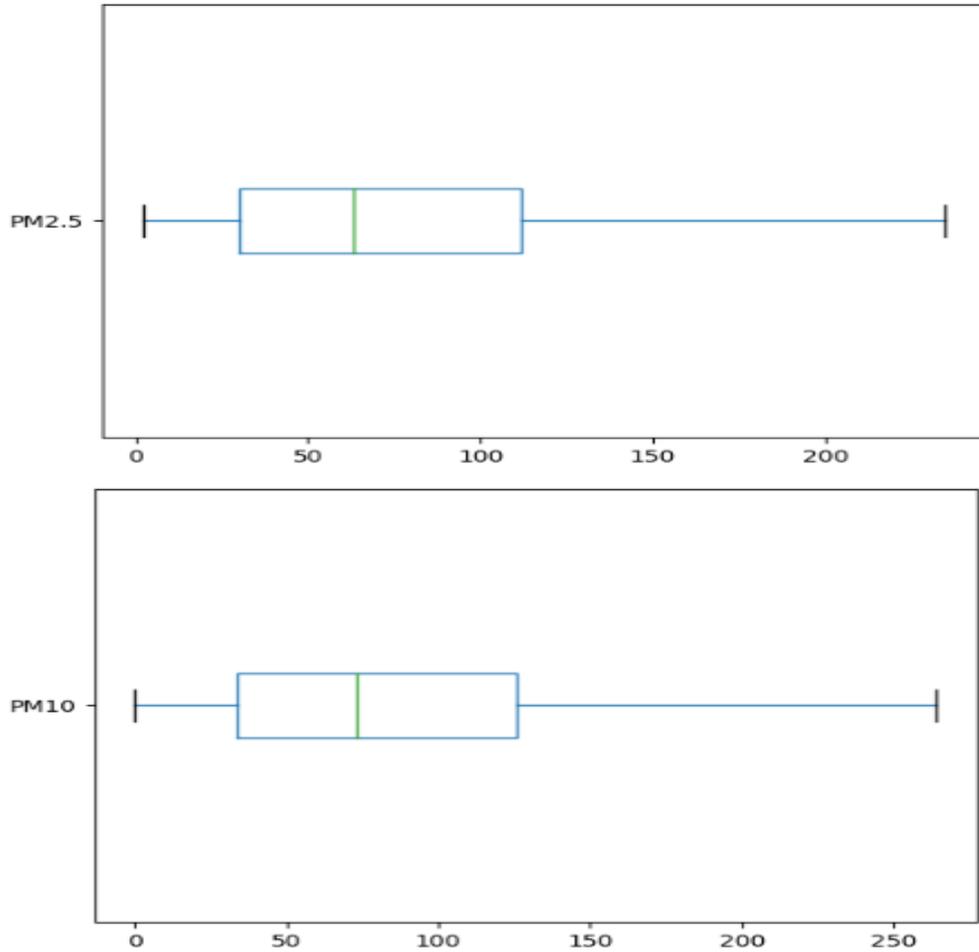


Figura 3.21 Gráfico de caja para los datos de $PM_{2.5}$ y PM_{10} tras la exclusión de valores atípicos.

Como lo muestra la Figura 3.21, las magnitudes de $PM_{2.5}$ Y PM_{10} , exhiben una distribución uniforme, dado que se han eliminado todas las observaciones atípicas que superaban su media (ver Figura 3.21).

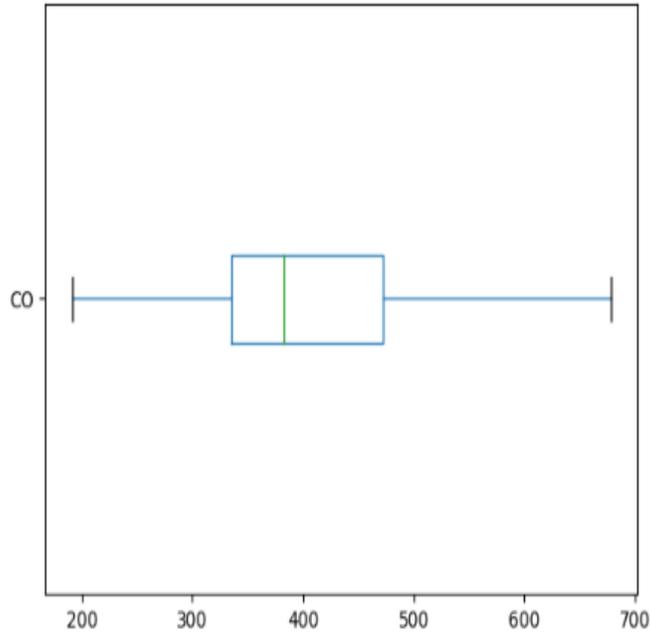


Figura 3.22 Grafico de caja de los datos de CO sin valores atípicos

Tal como se observa en la Figura 3.22, las mediciones de la magnitud CO presentan una distribución uniforme, dado que se procedió a eliminar todos los valores atípicos que se situaban por encima de su media (ver Figura 3.22).

Las mediciones de CO₂, temperatura y humedad muestran una distribución uniforme desde el inicio del proceso de detección de datos atípicos, ya que no se identificaron valores que se apartaran de su media. Esto se confirma mediante la observación de las Figuras 3.18, 3.19 y 3.20.

Los valores atípicos más extremos han sido significativamente reducidos y la exploración de las magnitudes continúa en curso, beneficiándose de la representación gráfica a través de los diagramas de caja y la normalización estandarizada. Es relevante destacar que el proceso de corrección de valores atípicos ha sido aplicado a todas las magnitudes numéricas presentadas en este documento.

3.5 Fase de análisis exploratorio y visualización de los datos

Los datos históricos de cada magnitud se examinan en las siguientes secciones de este documento para comprender su comportamiento y desarrollo histórico.

Para comprender dicho comportamiento, los datos históricos se dividen en los siguientes componentes:

- **Observed:** es la forma en la que se representan los datos.
- **Trend:** tendencia que se observa en los datos a través del tiempo.
- **Seasonal:** se refiere a los ciclos de repetición a través de la evolución de los datos.
- **Residual:** cambios en los datos a través de la tendencia.

3.5.1 Preparación de datos para representación gráfica

Como preparación previa a la generación de gráficos, se introducen nuevas columnas con el propósito de simplificar la visualización:

- **Year y Month:** estas columnas indican el año y mes de cada muestra. Estos datos resultan valiosos para identificar patrones recurrentes en diferentes meses a lo largo del año, lo que podría señalar comportamientos estacionales en los datos.

3.5.2 Exploración visual y extracción de la información

En el siguiente paso, se procede a la generación de representaciones gráficas individuales para cada una de las magnitudes, seguido de un análisis detallado de cada una de estas representaciones.

- $PM_{2.5}$: Tal y como se observa en la Figura 3.23, se hace evidente un patrón definido en el comportamiento del $PM_{2.5}$, evidenciando que los primeros meses del año presentan concentraciones notables de partículas de polvo. Esta tendencia se atribuye a las ráfagas de viento intensas durante dichos meses. Por otro lado, no se anticipa una variación sustancial en los meses posteriores, dado que estas partículas se generan consistentemente en concentraciones significativas (ver Figura 3.23).

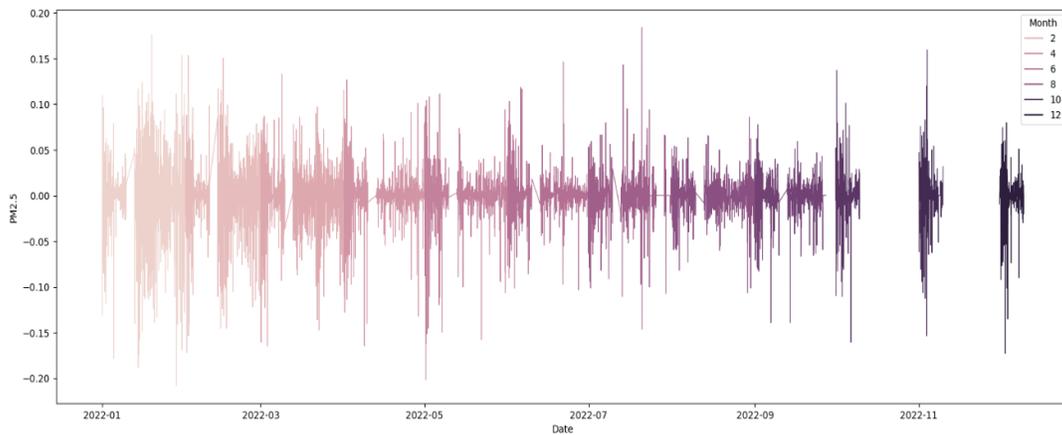


Figura 3.23 Datos históricos mensuales de $PM_{2.5}$

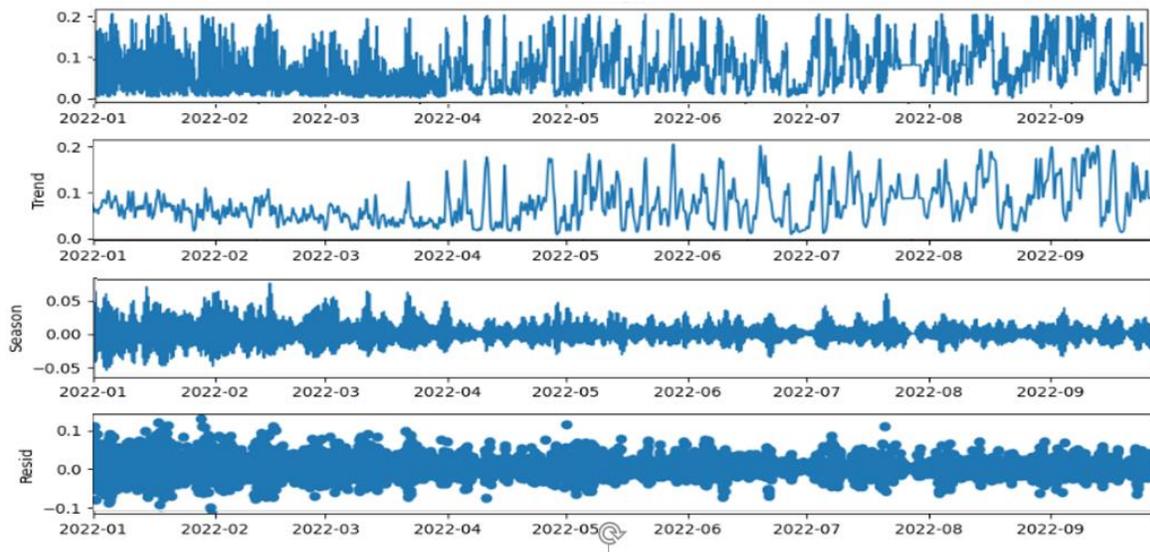


Figura 3.24 Descomposición estacional de $PM_{2.5}$

Tal y como se ilustra en la figura 3.24, se puede apreciar un incremento en las concentraciones de $PM_{2.5}$, desde enero hasta abril del año 2022, reflejando una tendencia estacional elevada. Sin embargo, se observa una disminución progresiva de estas concentraciones desde mayo hasta septiembre (ver Figura 3.24).

- PM_{10} : Tal como se detalla en el gráfico de la Figura 3.25, del mismo modo que con el $PM_{2.5}$, se evidencia que las concentraciones registradas mensualmente son notoriamente elevadas y exhiben escasa variabilidad (ver Figura 3.25). Esta constante alta concentración de partículas de polvo en el entorno es resultado, tanto de actividades humanas, como de procesos naturales, contribuyendo así a esta tendencia.

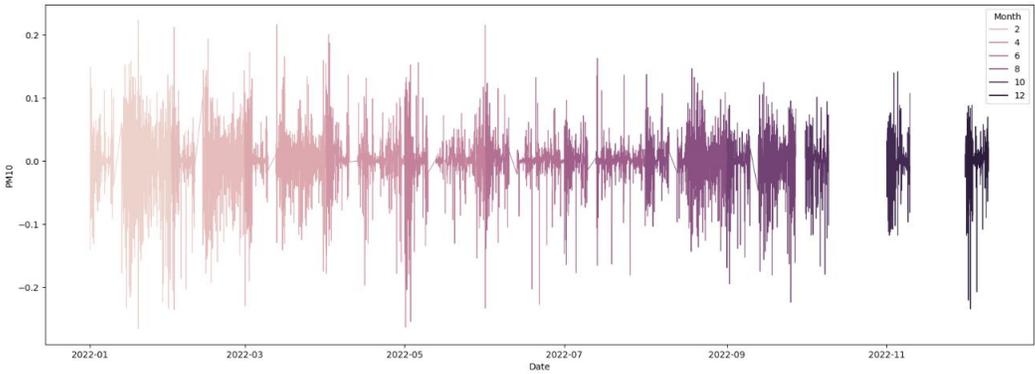


Figura 3.25 Datos históricos mensuales de PM_{10}

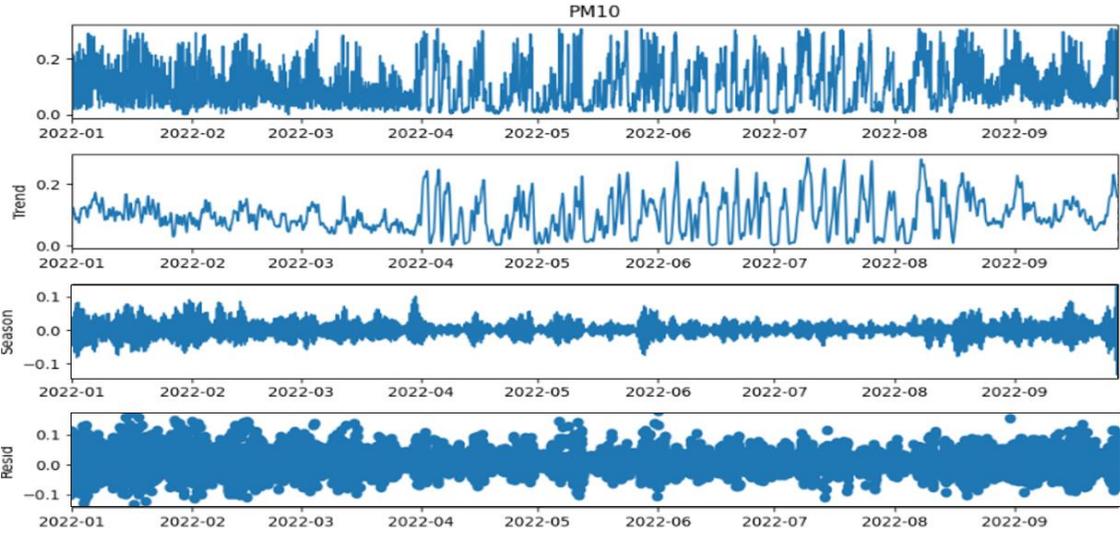


Figura 3.26 Descomposición estacional de PM_{10} .

La Figura 3.26 ilustra la estacionalidad de PM_{10} y revela cambios en las concentraciones que muestran una tendencia decreciente desde mayo hasta septiembre (ver Figura 3.26), seguida de un incremento en el mes de octubre.

- C_O : La Figura 3.27 muestra que los niveles de monóxido de carbono presentan un declive considerable a lo largo de los meses, mostrando una marcada variación estacional (ver Figura 3.27), Al someter su tendencia a un análisis similar al de la temperatura, se reafirma su propensión a la disminución.

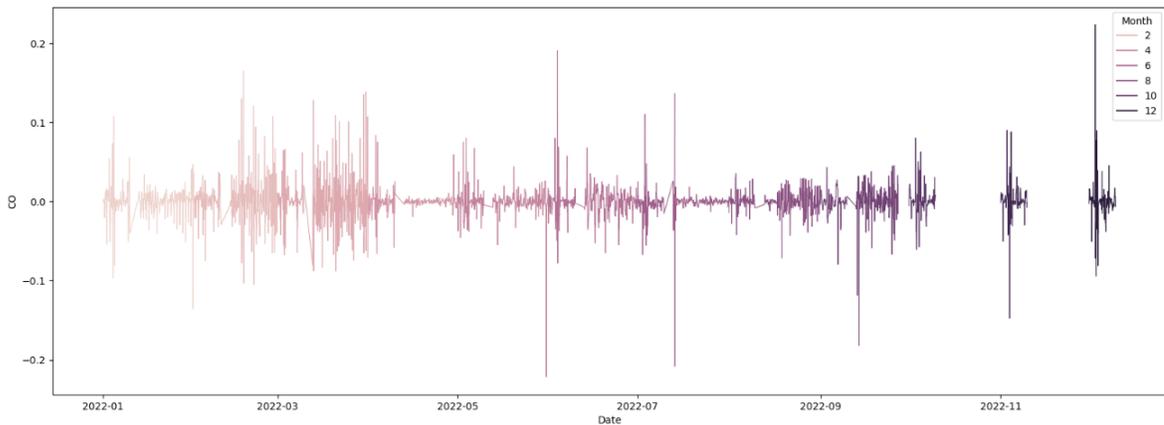


Figura 3.27 Datos históricos de CO

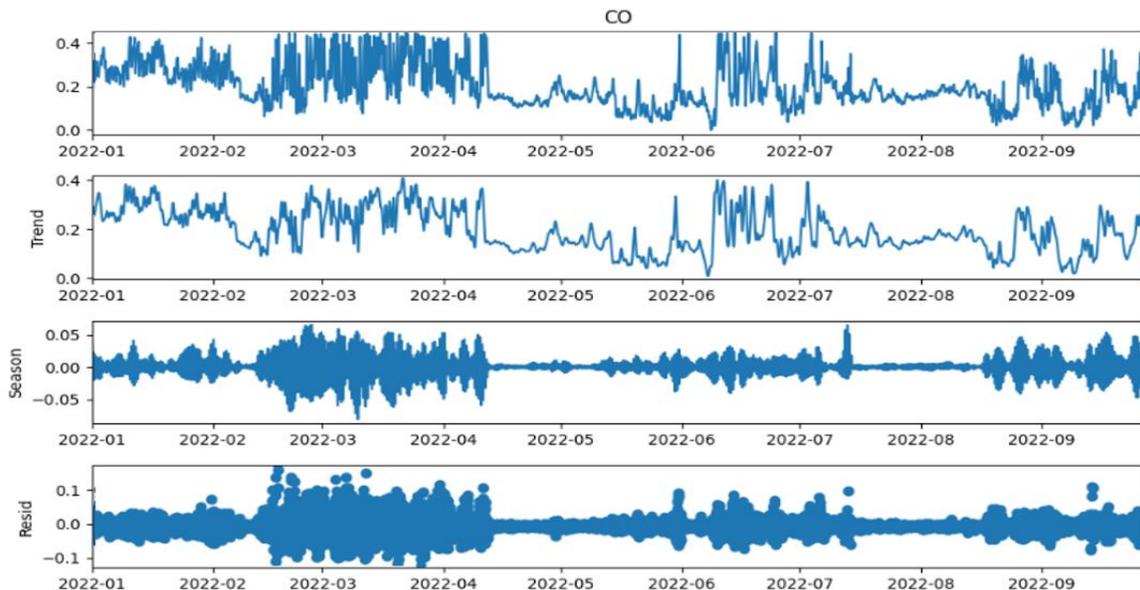


Figura 3.28 Descomposición estacional de CO.

La Figura 3.28 demuestra de manera clara una estacionalidad junto con una reducción en las emisiones de CO (ver Figura 3.28). Para validar esta observación estacional, se procede a evaluar el comportamiento de esta magnitud a lo largo de los meses correspondientes.

- CO_2 : De acuerdo con la gráfica de la Figura 3.29, los niveles de dióxido de carbono experimentan un notable aumento a medida que transcurren los meses, aunque no se evidencia una estacionalidad muy marcada (ver Figura 3.29).

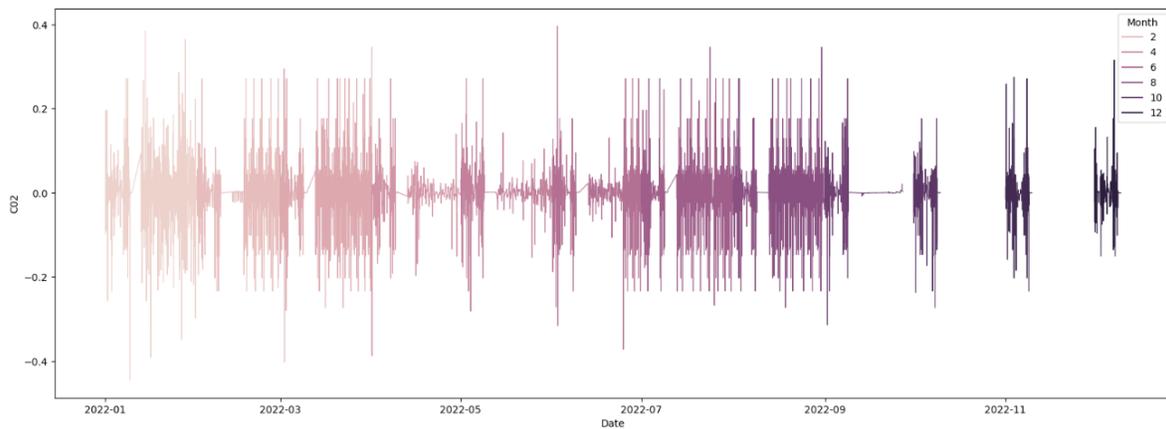


Figura 3.29 Datos históricos mensuales de CO_2

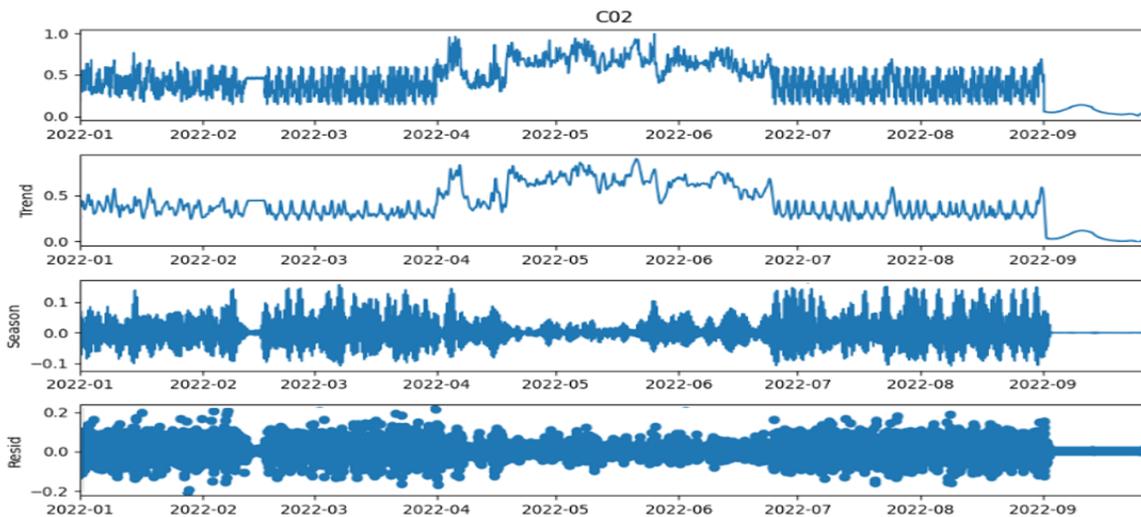


Figura 3.30 Descomposición estacional de CO_2

La representación gráfica en la Figura 3.30 muestra una estacionalidad, sin embargo, se nota una disminución apenas perceptible en las emisiones de dióxido de carbono en el transcurso de los meses (ver Figura 3.30).

- Temp: Al analizar la fluctuación de la temperatura media, se podría plantear como hipótesis inicial que los valores más elevados se manifestarán durante el verano, mientras que los más bajos se registrarán en invierno. Al explorar la tendencia que se observa en la Figura 3.31, a lo largo de los meses, es posible afirmar con un grado razonable de certeza que la temperatura sigue un patrón caracterizado por alcanzar valores mínimos al comienzo y al final del año, a la vez que alcanza valores máximos en el transcurso de la mitad del año (ver Figura 3.31).

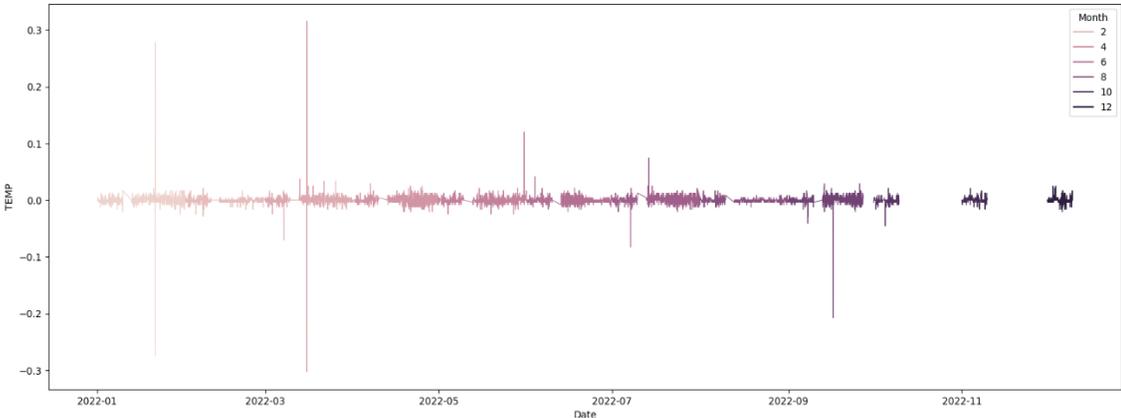


Figura 3.31 Datos históricos mensuales de Temperatura.

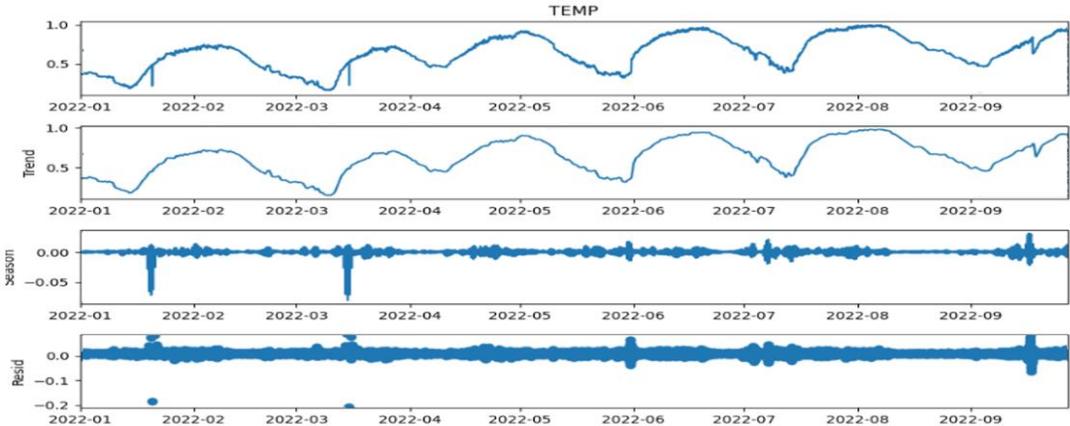


Figura 3.32 Descomposición estacional de temperatura.

Observando la tendencia en la Figura 3.32 se comprueba que los valores más altos de temperatura serán en verano y los más bajos en invierno (ver Figura 3.32).

- Hum: Al examinar la variabilidad de la humedad, se establece una hipótesis similar a la del estudio de la temperatura, anticipando que los valores más elevados se observarán durante el verano y los valores más bajos en invierno. Al estudiar la tendencia representada en la Figura 3.33, en función de los meses, se puede afirmar con cierta confianza que la humedad sigue un patrón distintivo, caracterizado por mostrar valores mínimos al inicio y al final del año, al mismo tiempo que exhibe valores máximos en el transcurso de la mitad del año (ver Figura 3.33).

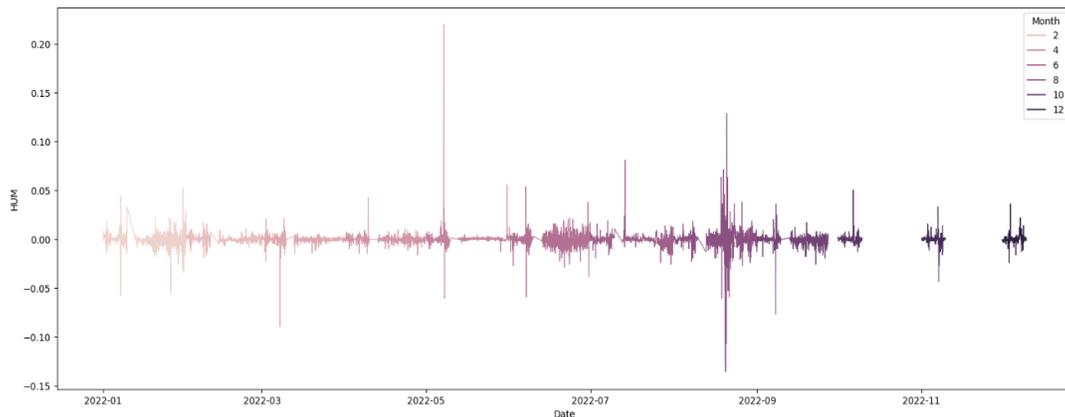


Figura 3.33 Datos históricos mensuales de humedad.

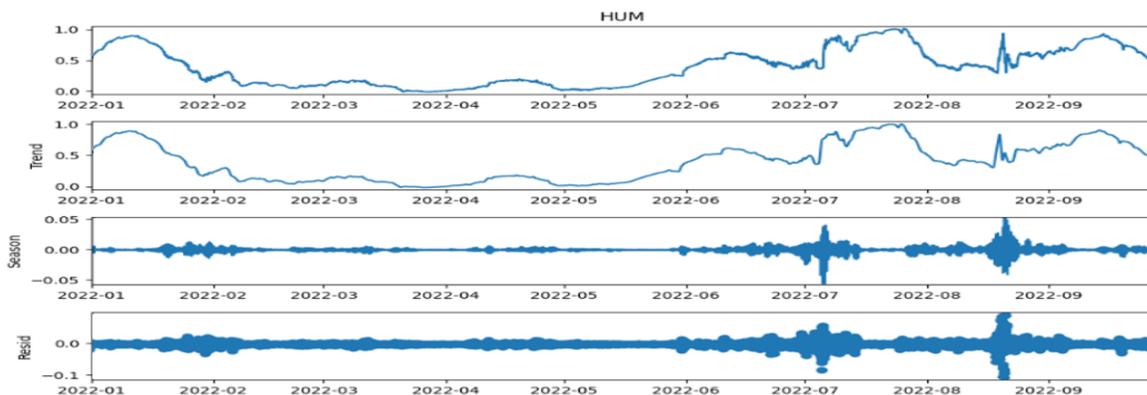


Figura 3.34 Descomposición estacional de humedad.

Observando la tendencia en la Figura 3.34 se comprueba que los valores más altos de humedad serán en verano y los más bajos en invierno (ver Figura 3.34).

3.5.3 Estimación de la densidad del kernel (KDE)

Estimar la función de densidad de probabilidad de la curva continua o no paramétrica del conjunto de datos en una o más dimensiones, permite crear un gráfico único para varias muestras, lo que ayuda a comprender de forma más eficiente la variación de los datos de las concentraciones de los contaminantes en el aire como lo muestran las siguientes figuras.

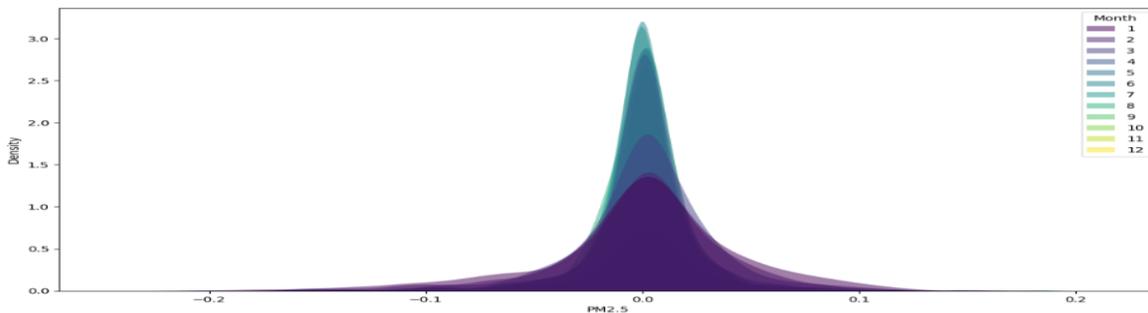


Figura 3.35 Estimación de densidad de $PM_{2.5}$.

Según la Figura 3.35, la densidad del $PM_{2.5}$ va subiendo con el paso de los meses y se estima que en los meses en donde subieron al máximo los niveles es en el mes de mayo y junio del año 2022 (ver Figura 3.35).

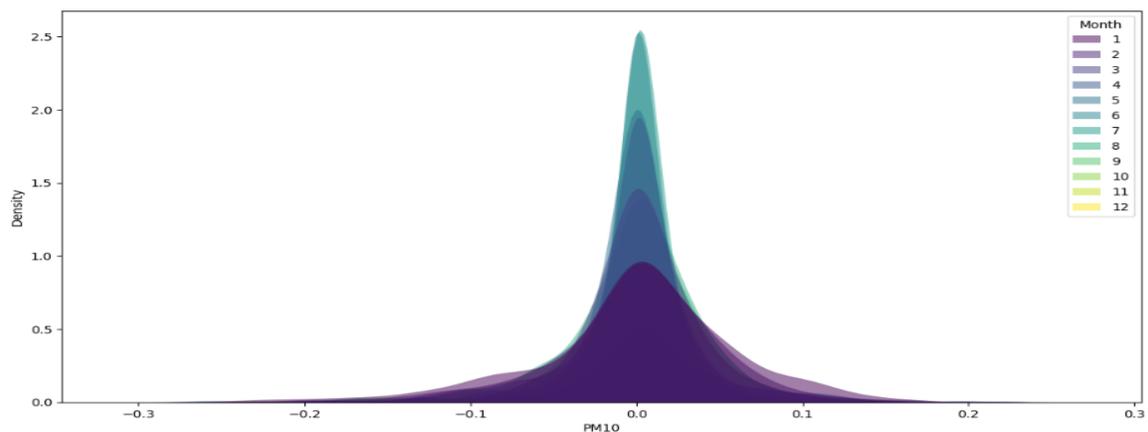


Figura 3.36 Diagrama de densidad de PM_{10} .

En la Figura 3.36, se puede observar que al igual que en el caso del $PM_{2.5}$ las concentraciones más altas de PM_{10} son en el mes de mayo y junio del año 2022 (ver Figura 3.36).

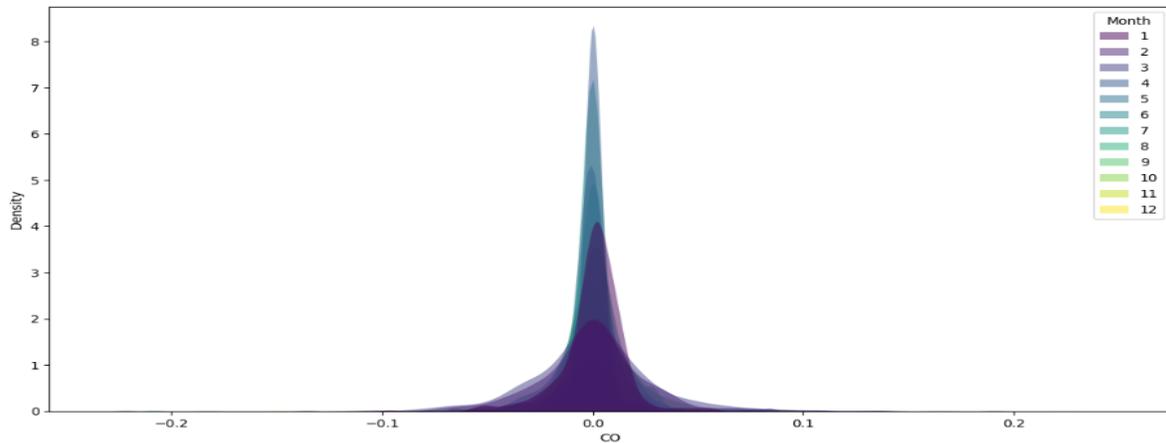


Figura 3.37 Diagrama de densidad de CO

En la Figura 3.37, se puede observar que, el monóxido de carbono tuvo concentraciones altas durante el mes de abril y mayo del año 2022 (ver Figura 3.37).

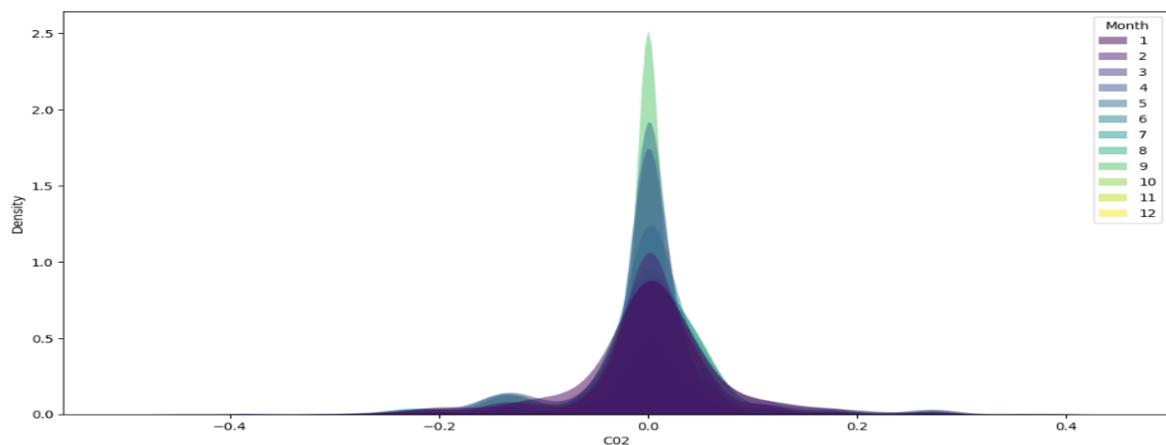


Figura 3.38 Diagrama de densidad de CO_2 .

En la Figura 3.38, se puede observar que, las concentraciones más altas de dióxido de carbono son en el mes de septiembre hasta diciembre, lo que quiere decir que, se elevaron debido al uso de calefacciones que usan algún tipo de gas natural (ver Figura 3.38).

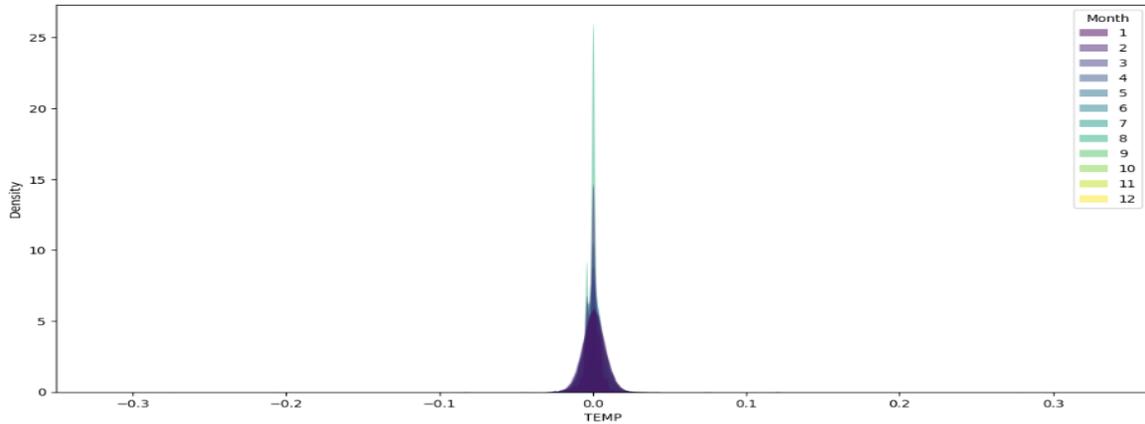


Figura 3.39 Diagrama de densidad de temperatura.

El diagrama de densidad mostrado en la Figura 3.39, indica las variaciones de temperatura durante los meses del año 2022. Siendo la más alta de 25 grados en el mes de agosto y la más baja de 5 grados en el mes de enero (ver Figura 3.39).

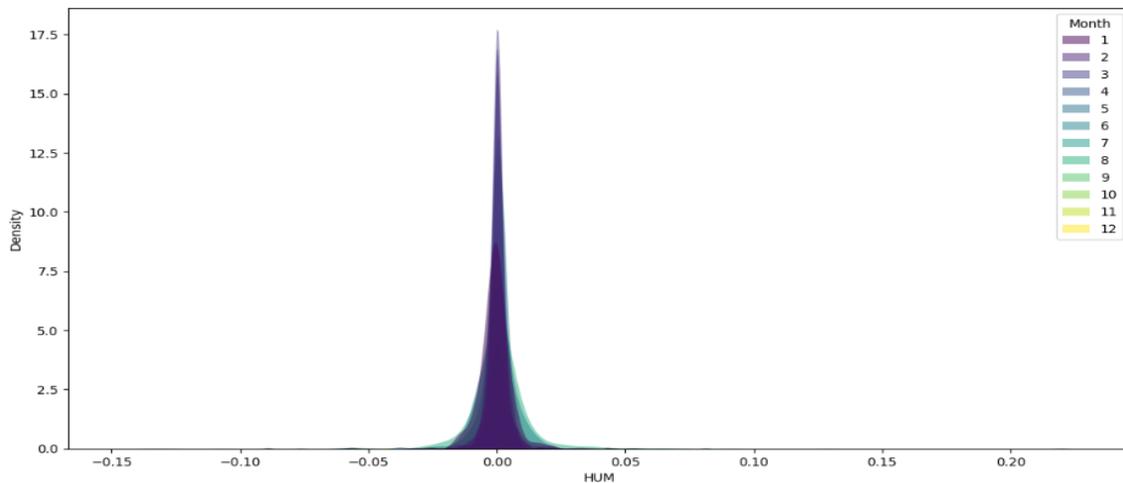


Figura 3.40 Diagrama de densidad de humedad.

El diagrama de densidad mostrado en la Figura 3.40, indica las variaciones de humedad durante los meses del año 2022. Partiendo del mes de enero a marzo se muestran niveles elevados de humedad, debido a que en estos meses se presentan bajas temperaturas y lluvias (ver Figura 3.40).

3.5.4 Diagramas de conjunto o dispersión

Para analizar la interacción entre las partículas de PM_{10} y otros contaminantes suspendidos en el aire interior, se recurre a la creación de gráficos de conjunto. Estos gráficos, generados mediante el uso de la librería Seaborn, presentan simultáneamente una relación bivariada y un perfil univariado. En esencia, se combina un diagrama de dispersión con un histograma para proporcionar una visualización completa.

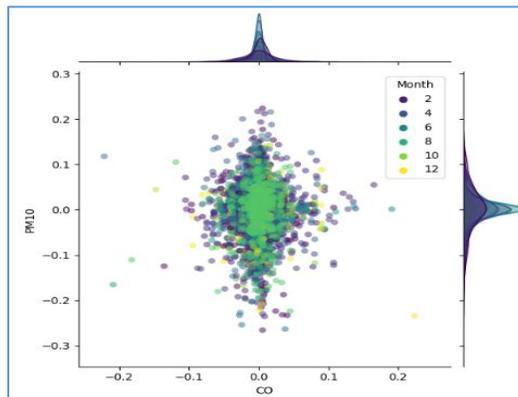


Figura 3.41 Diagrama de dispersión de PM_{10} con CO

Según la Figura 3.41, mientras haya mayor cantidad de monóxido de carbono en el aire menor será la cantidad de partículas de PM_{10} , (ver Figura 3.41).

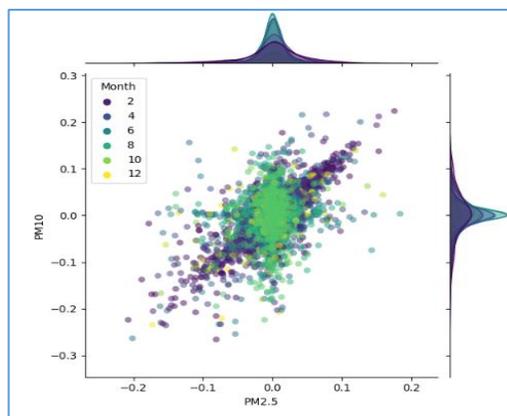


Figura 3.42 Diagrama de dispersión de PM_{10} con $PM_{2.5}$

Según el diagrama de la Figura 3.42, mientras mayor sea la cantidad de $PM_{2.5}$, los niveles de PM_{10} habrán disminuido (ver Figura 3.42).

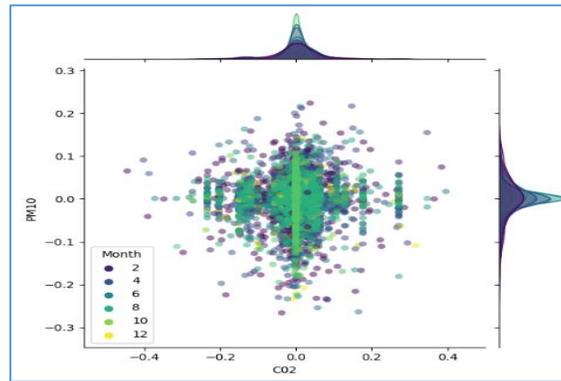


Figura 3.43 Diagrama de dispersión de PM_{10} con CO_2

Según la Figura 3.43, cuando haya concentraciones altas de dióxido de carbono en el aire, se encontrará muy poco PM_{10} , (ver Figura 3.43).

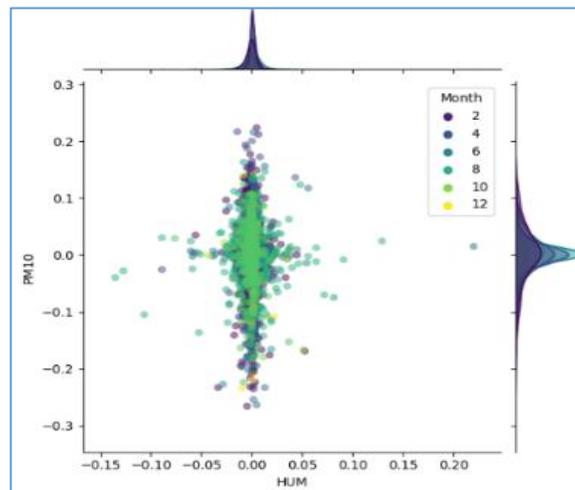


Figura 3.44 Diagrama de dispersión de PM_{10} con humedad

De acuerdo con la Figura 3.44, al igual que con la temperatura, aunque la humedad sea variante, siempre habrá niveles altos de PM_{10} en el aire interior, esto quiere decir que, entre las condiciones climáticas y las partículas de polvo siempre habrá relación (ver Figura 3.44).

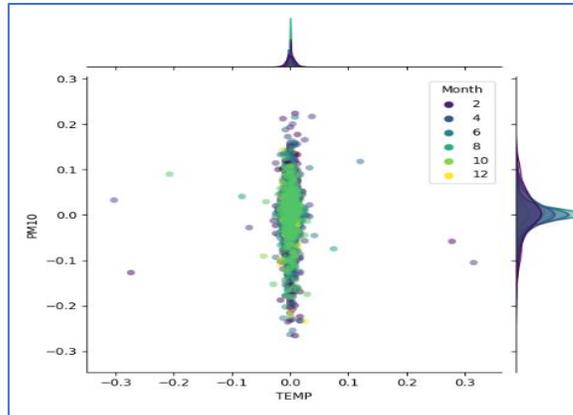


Figura 3.45 Diagrama de dispersión de PM_{10} con temperatura.

De acuerdo con la Figura 3.45, aunque las temperaturas sean variantes, siempre habrá niveles altos de PM_{10} en el aire interior, esto quiere decir que, entre las condiciones climáticas y las partículas de polvo siempre habrá relación (ver Figura 3.45).

3.5.5 Correlación lineal

La forma más sencilla de predecir futuros valores de una magnitud es estudiando su relación o correlación lineal con el resto de las variables. Esto muestra qué, tan fuerte sube (o baja) una magnitud cuando la otra cambia.

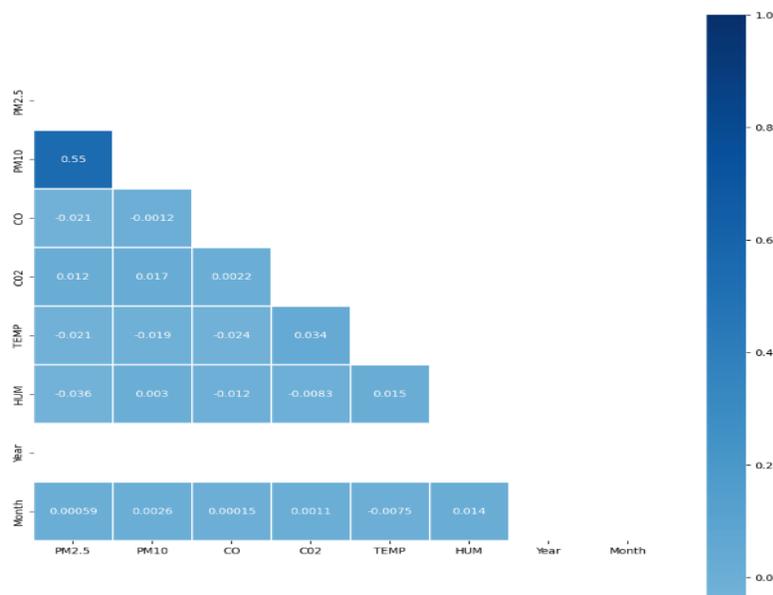


Figura 3.46 Matriz de correlaciones entre las distintas magnitudes.

Los valores de correlación máximos y mínimos son 1 y -1, respectivamente. Según la Figura 3.46, las magnitudes que más se relacionan entre sí son el CO y el CO₂, que tienen coeficientes de -0,21 y 0,012 con la temperatura y la humedad. Estos coeficientes muestran que una de las dos magnitudes tiende a disminuir ligeramente a medida que la otra aumenta (ver Figura 3.46).

3.5.6 Prueba de Dickey-Fuller aumentada

La prueba Dickey-Fuller aumentada (ADF) realiza una prueba de hipótesis para determinar si una serie de tiempo tiene una unidad raíz, es decir, si está definida por una tendencia donde se formulan dos hipótesis:

- H0 (hipótesis nula): los datos tienen unidad raíz por lo que no son estacionales.
- H1 (hipótesis alternativa): los datos no tienen unidad raíz así que son estacionales.

Si el valor p producido por la prueba es mayor que 0.05, se rechaza la hipótesis nula. En el caso de los datos sobre los contaminantes del aire interior, los hallazgos fueron los siguientes:

Tabla 3.7 Resultados Dickey-Fuller sobre los datos del año 2022 de los contaminantes suspendidos en el aire interior, así como sus condiciones climáticas.

Contaminante	P_VALOR
PM2.5	0.0
PM10	0.0
CO	0.0
CO2	0.0
Condiciones climáticas	
Temperatura	0.0
Humedad	0.0

Dado que el valor de p en todas las variaciones que se encuentran en la Tabla 3.7, es menor que 0.05 se rechaza la hipótesis nula y se comprueba que los datos son

estacionales y por lo tanto se puede proceder a aplicar el modelo ARIMA para hacer predicciones (ver Tabla 3.37).

La auto-correlación y correlación parcial de los datos de contaminantes, a través del número de rezagos, se puede utilizar para estimar manualmente los parámetros del modelo ARIMA. Para probar un rango de valores para cada parámetro y seleccionar el que mejor se ajuste a los datos, se utilizará una función para encontrar los valores más adecuados.

Se han creado dos grupos del conjunto de datos:

- Datos de entrenamiento: un conjunto de datos que se utiliza para calibrar el modelo ARIMA.
- Datos de prueba: son el conjunto de datos sobre el cual se compararán las predicciones y se calcularán los errores.

Debe tener en cuenta que no habrá una predicción para tantos valores como valores de prueba; más bien, habrá tantas predicciones para el siguiente valor como datos de prueba. En otras palabras, no se realizará una predicción de X valores futuros; en su lugar, se realizarán X predicciones de un solo valor. El modelo se actualizará con los datos de entrenamiento y el valor de prueba correspondiente para cada predicción. El objetivo de este modo de predicción es minimizar el error, tanto como sea posible. La Tabla 3.8 muestra los parámetros del modelo ARIMA que mejor se ajustan a los datos de entrenamiento.

Tabla 3.8 modelos ARIMA definitivos que se escogieron de acuerdo con los criterios de ajuste.

Contaminante	ARIMA
PM2.5	p: 2,d: 0, q: 1
PM10	p: 2,d: 0, q: 1
CO	p: 0,d: 1, q: 1
CO2	p: 2,d: 0, q: 1

En la Tabla 3.8 se muestran los modelos ARIMA definitivos que se escogieron, de acuerdo con los criterios de ajuste. De esta manera, se observó que la mayoría

describieron una estructura temporal similar para tres contaminantes, a excepción de uno, al cual se le ajustó un modelo distinto al del resto (ver Tabla 3.8).

Para empezar a abordar el análisis de los resultados arrojados entre los contaminantes del aire, se observaron las predicciones mensuales de las fluctuaciones que presentaron todos los contaminantes en el aire interior.

- Escalas mensuales de $PM_{2.5}$ Y PM_{10} :

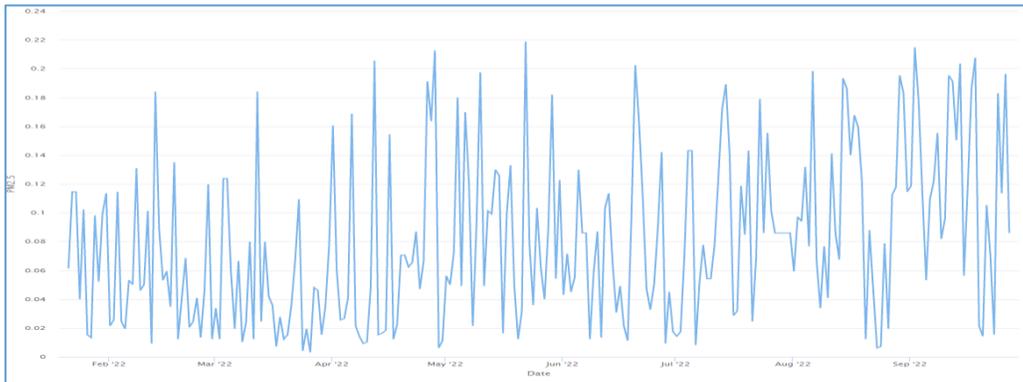


Figura 3.47 Predicción mensual de $PM_{2.5}$

En la Figura 3.47 se observa que las predicciones a nivel mensual indicaron que las partículas de $PM_{2.5}$ tuvieron una permanencia de manera uniforme, encontrando las mismas variaciones que se muestran en los picos de los datos de entrenamiento (ver Figura 3.47).

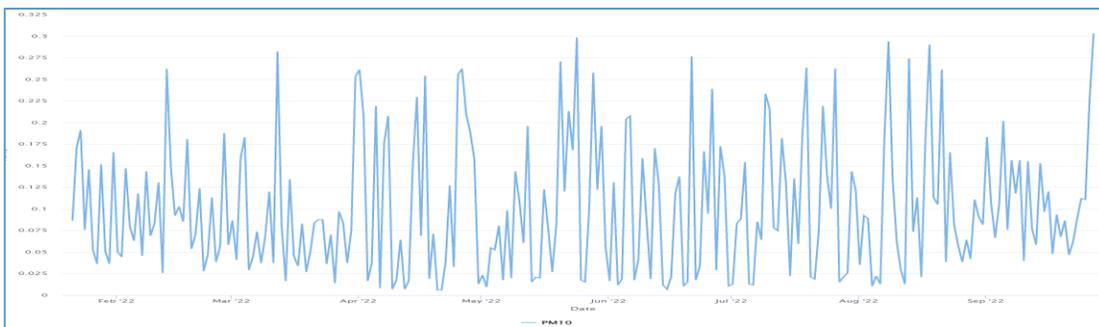


Figura 3.48 Predicción mensual de PM_{10}

En la Figura 3.48, se observan las predicciones a nivel mensual de PM_{10} , las cuales indicaron que hay una permanencia de manera uniforme en todos los meses en el año, lo cual se ve reflejado en la similitud que tienen los picos de los datos de prueba con los datos de entrenamiento (ver Figura 3.48).

- Escala de tiempo mensual de CO Y CO_2 :

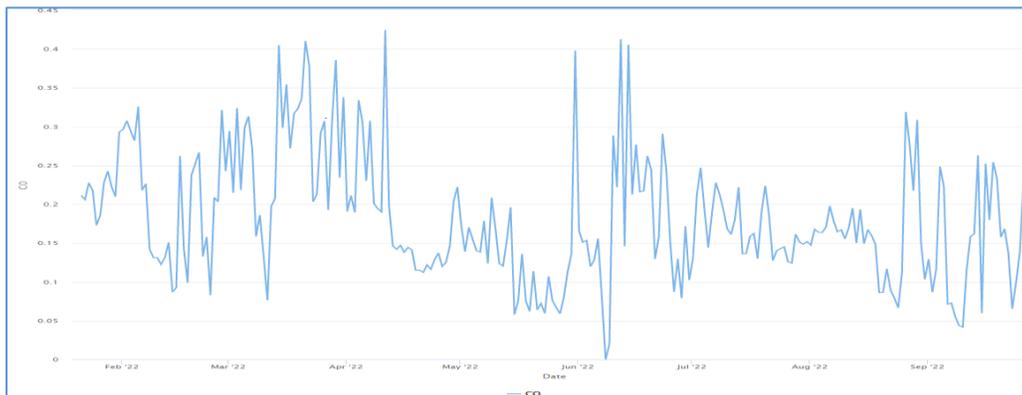


Figura 3.49 Predicción mensual de CO .

En la Figura 3.49, se observan las predicciones a nivel mensual de monóxido de carbono, las cuales indican una permanencia de manera uniforme en todos los meses en el año, demostrando que las concentraciones más fuertes de este contaminante se presentaron en los picos de los datos de abril a junio (ver Figura 3.49).

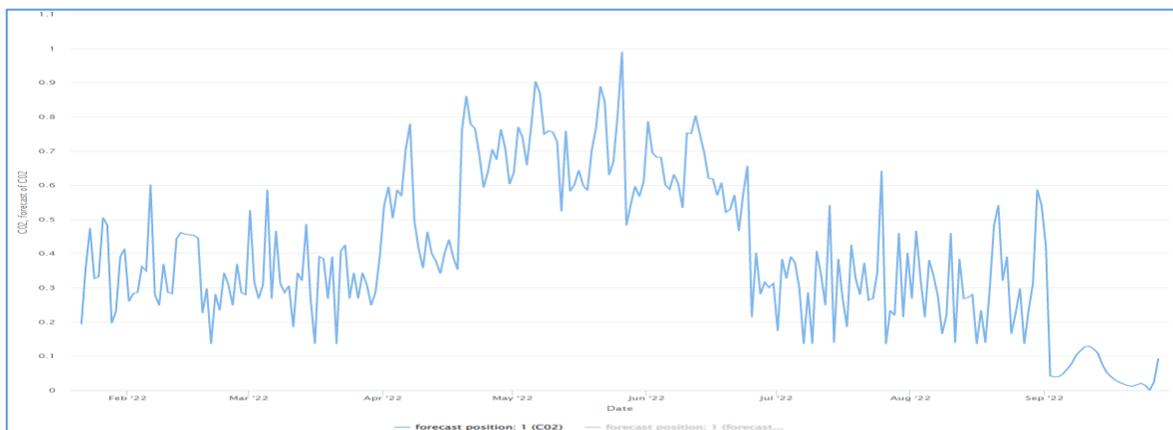


Figura 3.50 Predicción mensual de CO_2

En la Figura 3.50, se observan las predicciones a nivel mensual de dióxido de carbono, las cuales indicaron que hay una permanencia de manera uniforme en todos los meses en el año, lo cual demuestra que las concentraciones más altas de este contaminante se centran en los meses de abril a mediados de junio, al igual que con el monóxido de carbono (ver Figura 3.50).

3.5.7 Árboles de decisión potenciados por gradientes

- Árbol de decisión potenciado por gradientes de $PM_{2.5}$ Y PM_{10} :

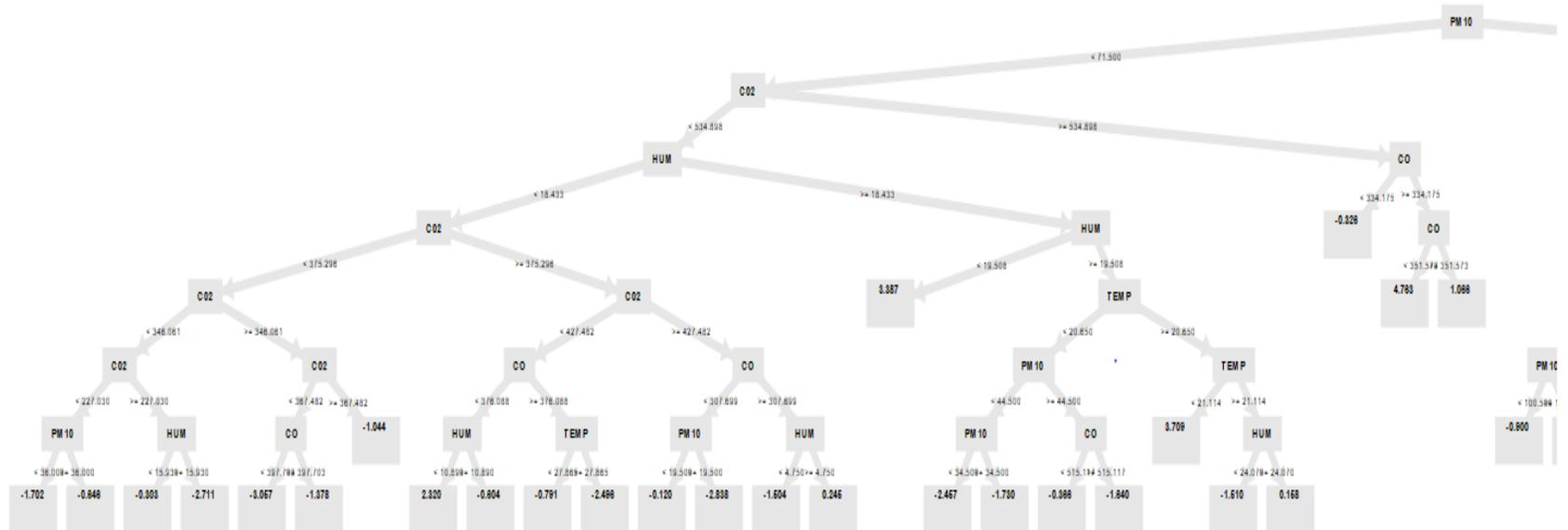


Figura 3.51 Árbol de decisión potenciado por gradientes de PM_{10} y $PM_{2.5}$ parte 1

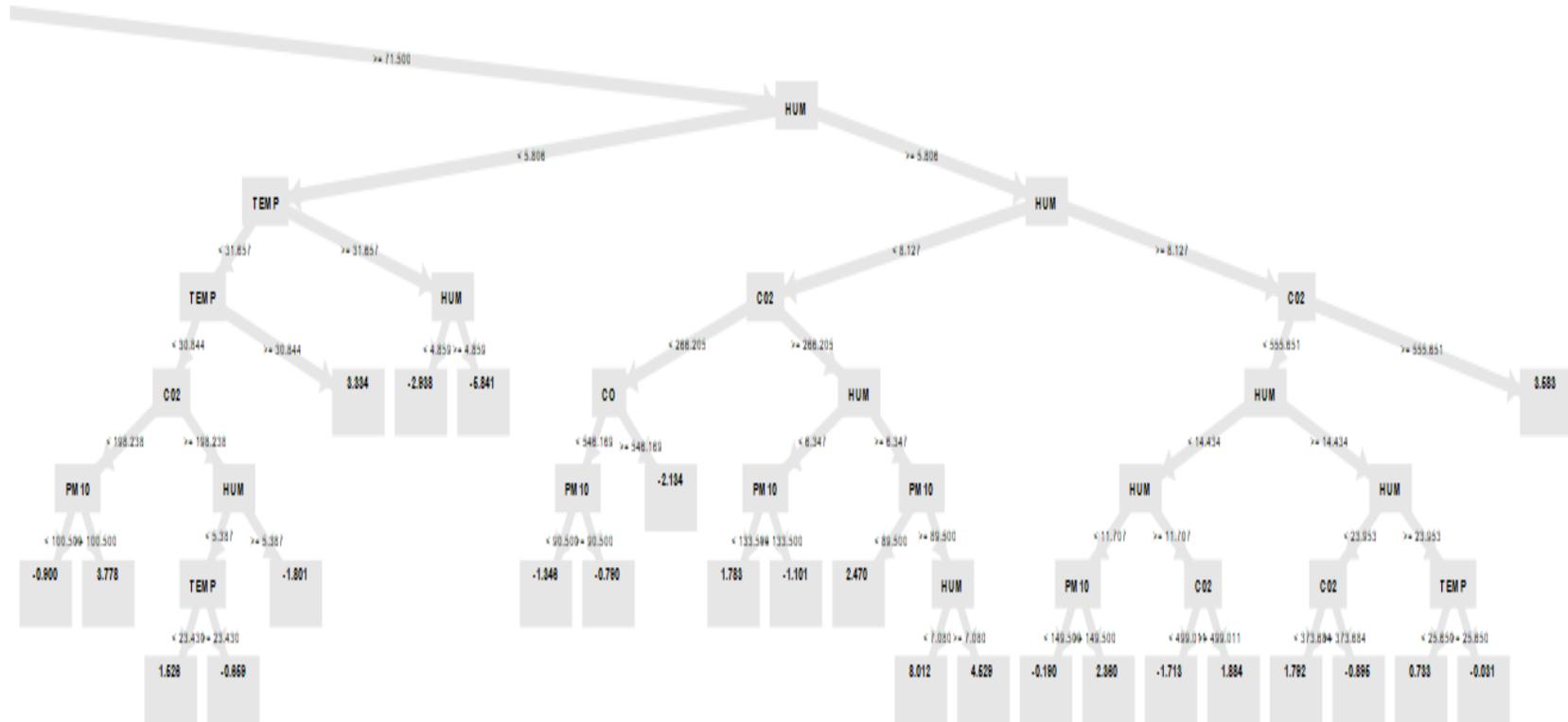


Figura 3.52 Árbol de decisión potenciado por gradientes de PM_{10} y $PM_{2.5}$ parte 2

El árbol de decisiones de la Figura 3.51 y Figura 3.52, nos ayuda a optimizar el valor predictivo del modelo a través de pasos sucesivos en el proceso de aprendizaje. Cada iteración del árbol de decisión implica ajustar los valores de los coeficientes, pesos o sesgos aplicados a cada una de las variables de entrada que se utilizan para predecir el valor objetivo, buscando minimizar la función de pérdida, es decir, la medida de la diferencia entre el valor previsto y el valor objetivo real. El gradiente es el ajuste incremental realizado en cada paso del proceso; *boosting* es un método para acelerar la mejora en la precisión predictiva a un valor suficientemente óptimo (ver Figuras 3.51 y Figura 3.52).

Según la predicción del árbol de decisión se puede encontrar que:

- Cuando se encuentren concentraciones bajas de PM_{10} , el dióxido de carbono habrá aumentado sus niveles.
- Al encontrarse concentraciones altas de dióxido de carbono, la humedad habrá aumentado.
- Cuando las concentraciones de dióxido de carbono se encuentren bajas, será porque la humedad y el PM_{10} en el ambiente habrán aumentado.
- Al haber concentraciones altas de dióxido de carbono, se encontrarán altas las concentraciones de humedad.
- Al aumentar la humedad y la temperatura, se encontrarán concentraciones bajas de PM_{10} , pero altas de monóxido de carbono.
- Al bajar la temperatura, se encontrarán concentraciones altas de dióxido y monóxido de carbono.

- Árbol de decisión potenciado por gradientes de CO

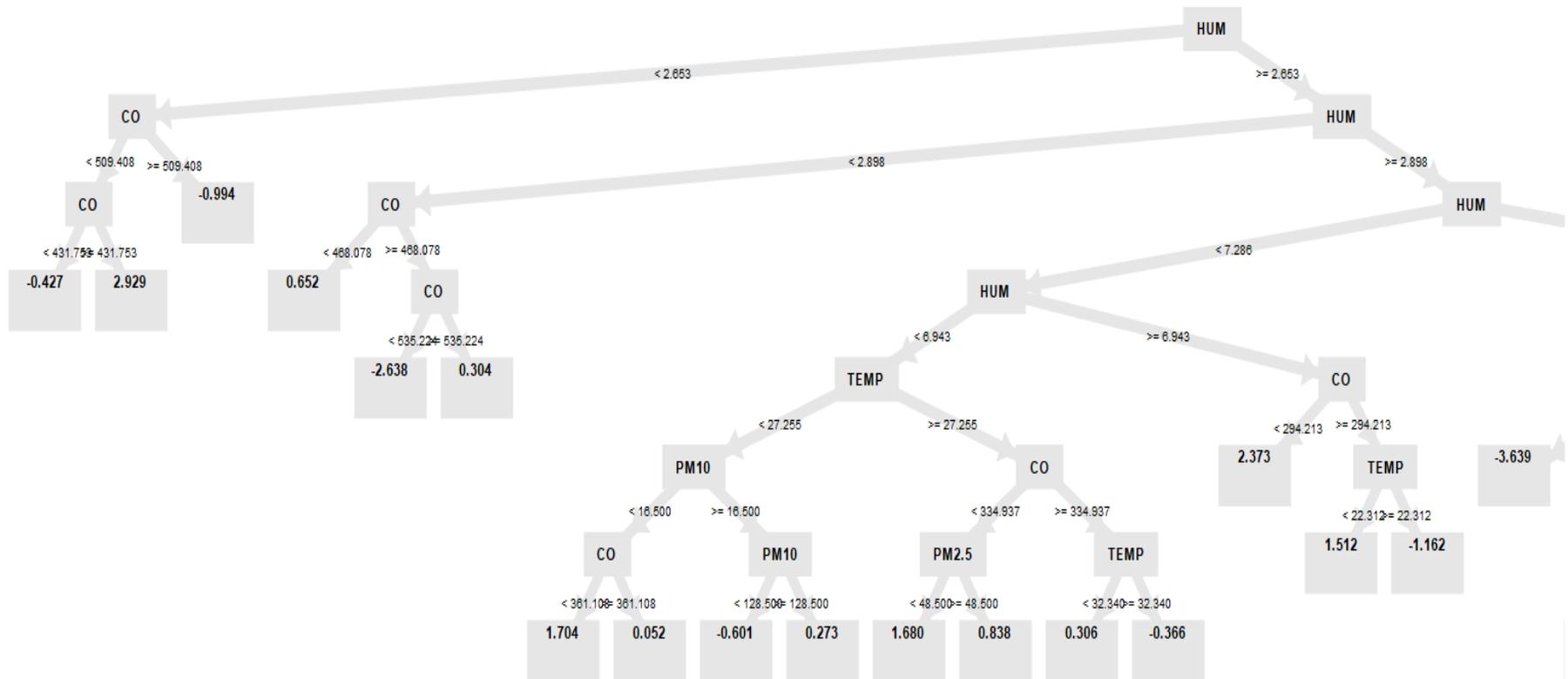


Figura 3.53 Árbol de decisión potenciado por gradientes de CO parte 1.

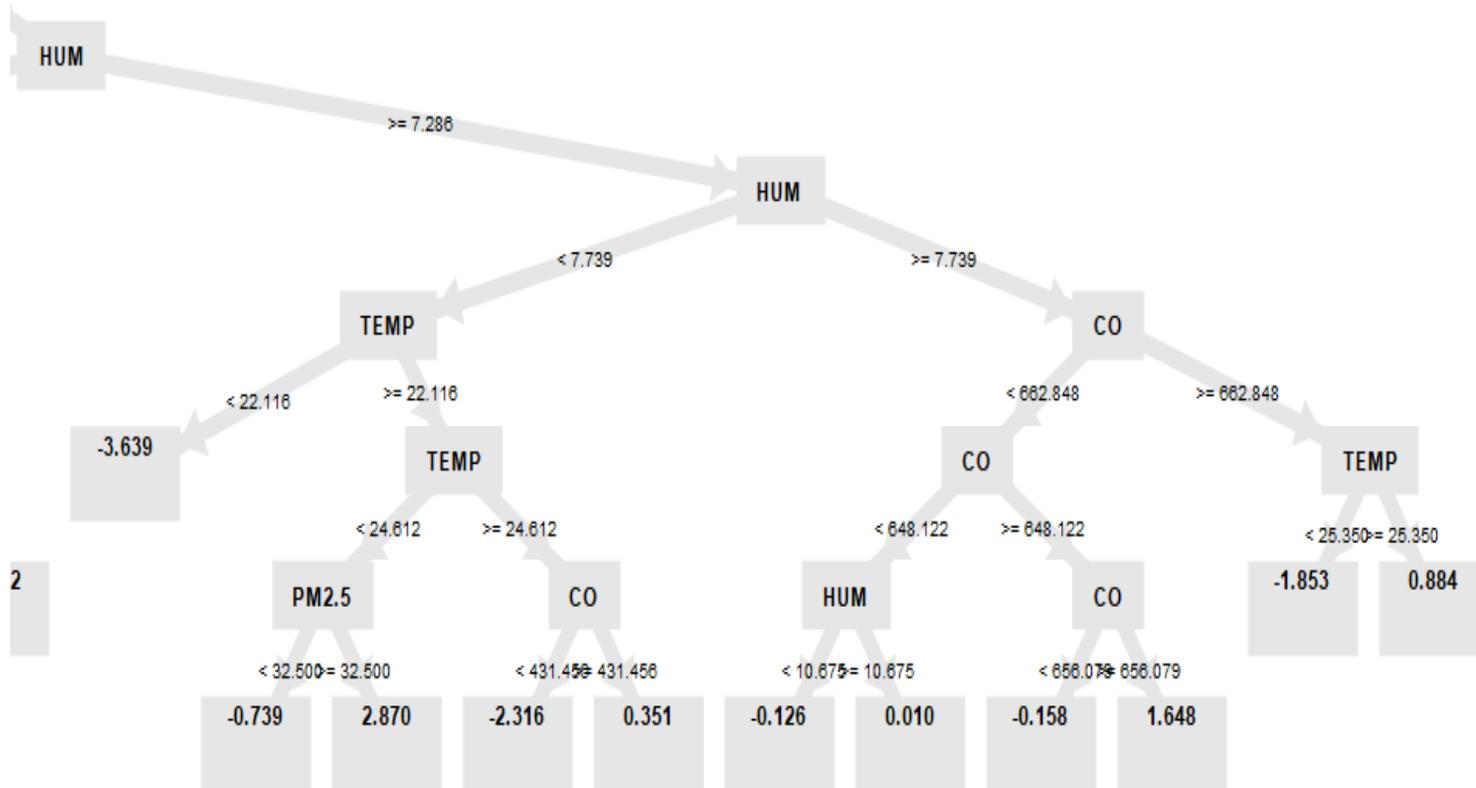


Figura 3.54 Árbol de decisión potenciado por gradientes de CO parte 2

Según la predicción del árbol de decisión de las Figuras 3.53 y Figura 3.54, se puede observar que:

- Cuando la temperatura sea menor y el valor de la humedad sea mayor, el nivel de CO habrá aumentado.
- Si la humedad es menor y la temperatura mayor, se encontrarán partículas de CO₂.
- Al disminuir el valor de la temperatura y la humedad, se encontrarán partículas de PM_{2.5} y PM₁₀ en el aire.
- Si las concentraciones de PM₁₀ y PM_{2.5} son bajas, se encontrarán partículas de CO elevadas.
- Cuando la humedad sea mayor que los demás contaminantes, se encontrará CO en concentraciones elevadas (ver Figuras 3.53 y 3.54).

- Árbol de decisión potenciado por gradientes de CO_2

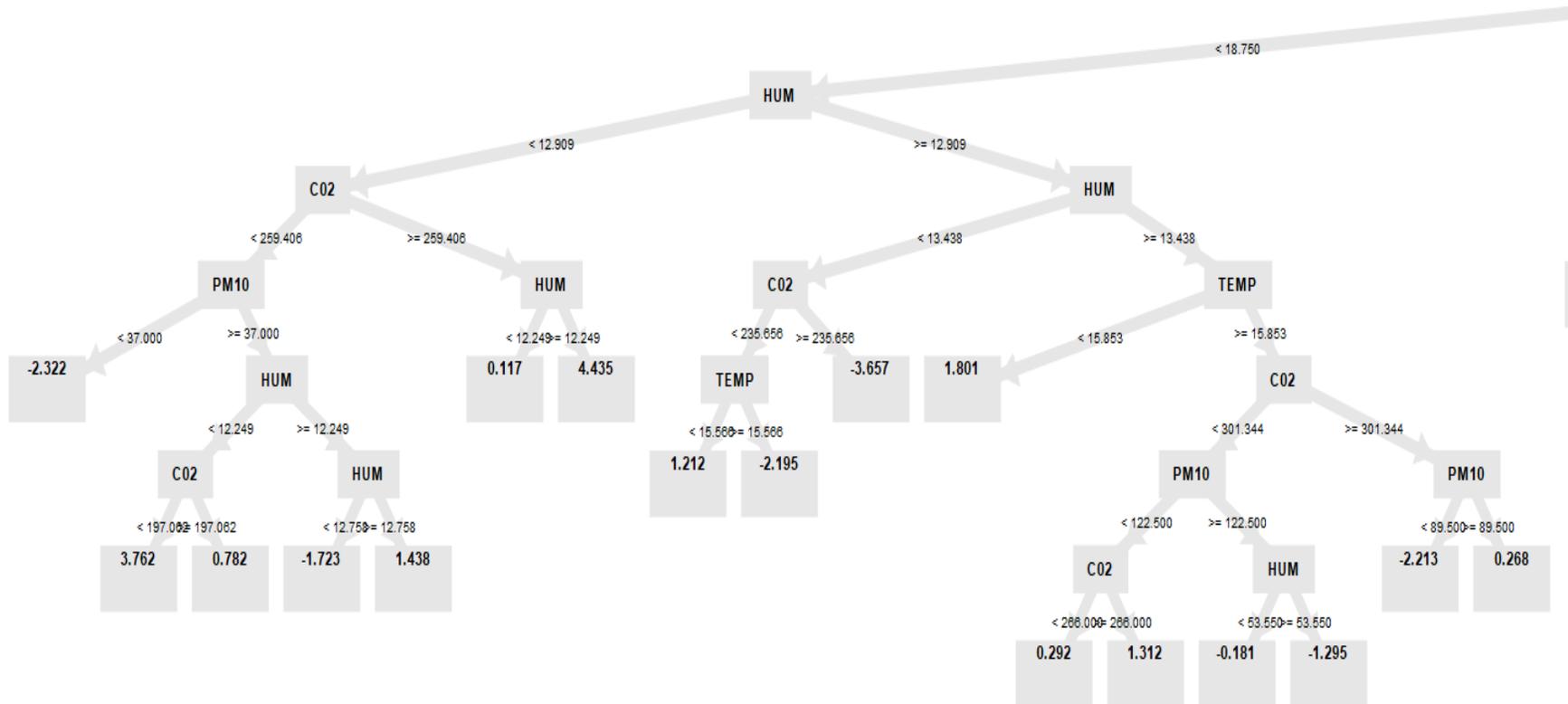


Figura 3.55 Árbol de decisión potenciado por gradientes de CO_2 parte 1.

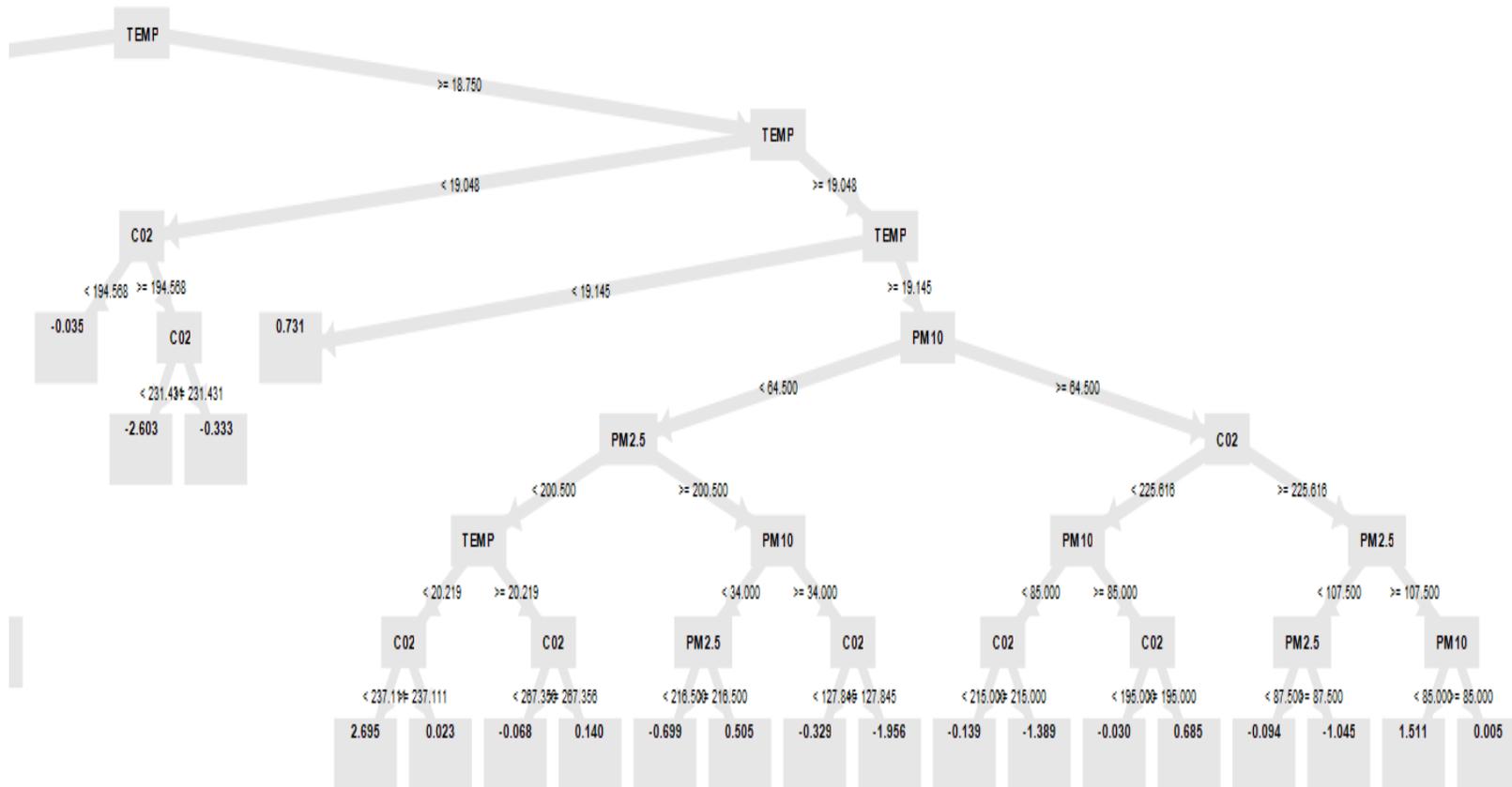


Figura 3.56 Árbol de decisión potenciado por gradientes de CO₂ parte 2

De acuerdo con la predicción del árbol de decisión de las Figuras 3.55 y Figura 3.56, se puede observar lo siguiente:

- Si las condiciones climáticas de humedad y temperatura son bajas, se presentarán concentraciones elevadas de PM_{10} y $PM_{2.5}$.
- Cuando las concentraciones de CO_2 sean bajas, se presentarán niveles elevados de PM_{10} y $PM_{2.5}$.
- Si las concentraciones de PM_{10} y $PM_{2.5}$ son bajas y la humedad es mayor, habrán emisiones altas de CO_2 .
- Cuando la temperatura sea baja y la humedad sea alta, las concentraciones de CO serán elevadas.
- Cuando el valor del PM_{10} sea bajo habrá aumentado el $PM_{2.5}$.
- Mientras el valor de la temperatura sea mayor a los contaminantes, las emisiones de CO_2 se encontrarán en aumento (ver Figuras 3.55 y figura 3.56).

3.5.8 Correlación generada por los árboles de decisión

La Tabla 3.9 muestra las correlaciones resultantes de las predicciones hechas por los árboles de decisión potenciados por gradientes.

Tabla 3.9 Correlación de variables de acuerdo con árboles de decisión.

Attributes	C02	CO	HUM	PM10	PM2.5	TEMP
C02	1	-0.074	-0.389	-0.065	-0.052	0.039
CO	-0.074	1	-0.066	-0.027	-0.118	-0.230
HUM	-0.389	-0.066	1	0.180	0.210	0.070
PM10	-0.065	-0.027	0.180	1	0.242	-0.038
PM2.5	-0.052	-0.118	0.210	0.242	1	0.121
TEMP	0.039	-0.230	0.070	-0.038	0.121	1

En la tabla 3.9, se puede observar una correlación más clara, que nos dice que las emisiones de dióxido y monóxido de carbono están relacionadas con las condiciones climáticas, mientras que el material particulado, está poco relacionado con las otras variables (ver Tabla 3.9).

CAPÍTULO IV. RESULTADOS OBTENIDOS

4.1 Funcionamiento del prototipo

Al exponer el prototipo del módulo de la red de contaminantes del aire interior a un ambiente real, se pudo observar que adquirió un comportamiento similar a la clasificación de las diferentes escalas del ICA, un índice para la notificación diaria de la calidad del aire, que va en una escala de 0 a 500 y que señala el grado de pureza o contaminación atmosférica de nuestro territorio.

4.2 Análisis inicial de los datos

Dado que los problemas discutidos en la sección 1.1 son donde se contextualiza la importancia de los contaminantes atmosféricos, es posible interpretar tecnológicamente el análisis de datos históricos, desde un punto de vista centrado en esos temas.

Las condiciones atmosféricas se mantienen estables a través de todo el marco de tiempo contemplado en este estudio, si bien sería necesario realizar un estudio mucho más amplio en el tiempo para sacar otro tipo de conclusiones. Por otro lado, una gran mayoría de los contaminantes han visto sus niveles reducidos en el último año, en especial el CO y CO_2 han sufrido una reducción muy pronunciada. Sin embargo, los niveles de $PM_{2.5}$ y PM_{10} , en vez de disminuir, se mantienen elevados, debido a que siempre hay mucha actividad humana dentro de los ambientes interiores.

4.3 Análisis de Auto-regresión con modelo ARIMA

En el enfoque principal para predecir los valores futuros de la contaminación del aire interior, ha sido utilizando el modelo ARIMA para captar el comportamiento puramente estacional de las series temporales.

Al hacer las predicciones mensuales futuras de los contaminantes que se encontraron en el aire interior, se obtuvo la Tabla 4.1 con los resultados de los márgenes de error.

Tabla 4.1 Resultados de las predicciones futuras de los contaminantes del aire interior.

Predicción de:	Error cuadrático medio (RMSE)	Error absoluto medio (MAE)
PM_{2.5}	0.017	45.97%
PM₁₀	0.025	39.78%
CO	0.011	5.85%
CO₂	0.034	13.63%

Como se observa en la Tabla 4.1, al comparar los resultados de los modelos ARIMA hechos para la predicción de los contaminantes del aire, se puede notar que el que tiene un error menor a los demás contaminantes es el del monóxido de carbono, ya que como se observó en la sección 3.5.6, se tuvo que ajustar a parámetros diferentes para que la predicción fuera estable, aunque hay que recordar que los datos sobre los que se ha ajustado el modelo son los correspondientes a un solo año (ver Tabla 4.1).

4.4 Resultados datos reales vs ARIMA

A continuación, se mostrarán los resultados de las predicciones ARIMA versus los datos reales representados en una gráfica de líneas.

- Resultados de las predicciones de $PM_{2.5}$

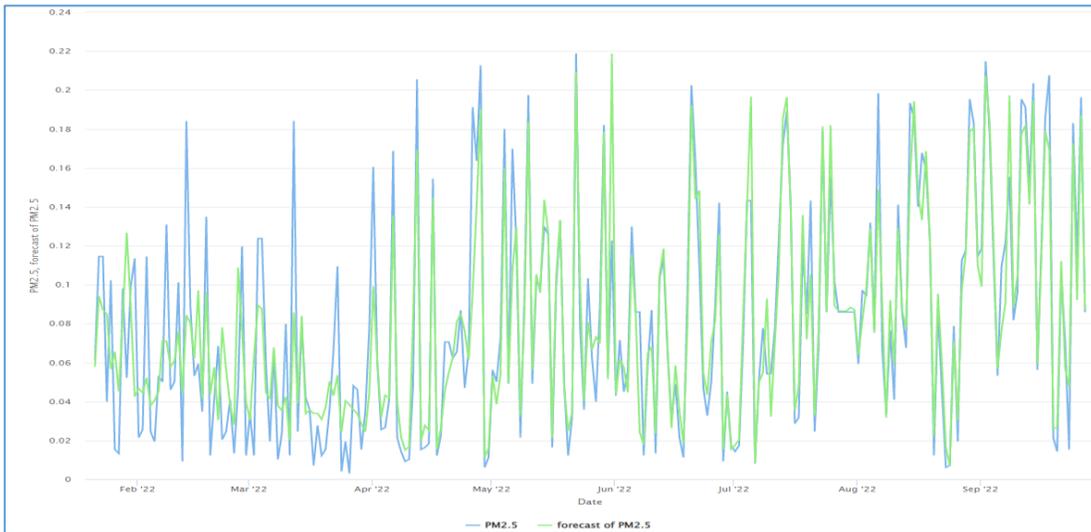


Figura 4.1 Comparación entre datos reales de $PM_{2.5}$ y el método ARIMA

En la Figura 4.1, se puede ver que los datos actuales, marcados de color azul en el gráfico y los datos de la predicción del modelo ARIMA, marcados de color verde, cuentan con bastante similitud (ver Figura 4.1), lo que quiere decir que se pudo hacer una buena estimación de los datos a futuro, a pesar de su margen de error del 45.97%.

- Resultados de las predicciones de PM_{10}

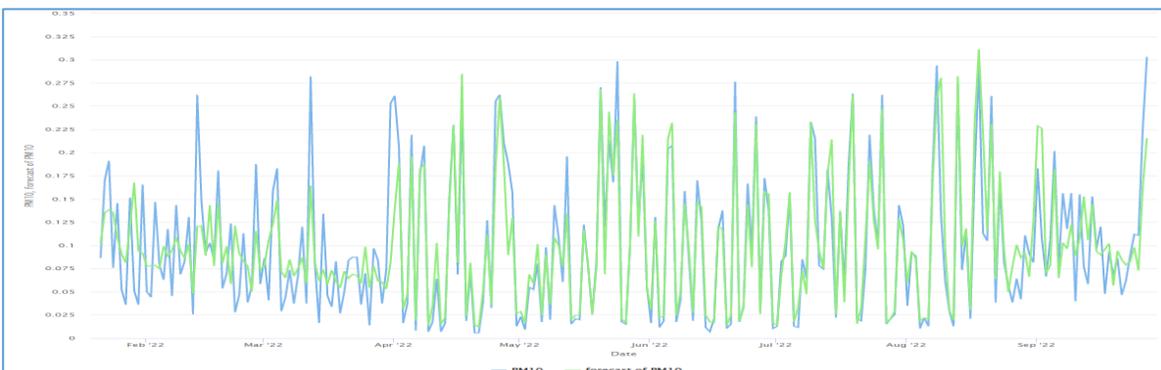


Figura 4.2 Comparación entre datos reales de PM_{10} y el método ARIMA

En la Figura 4.2, se puede ver que los datos actuales de PM_{10} , marcados de color azul en el gráfico y los datos de la predicción del modelo ARIMA, marcados de color verde, cuentan con bastante similitud, lo que quiere decir que se pudo hacer una buena estimación de los datos a futuro (ver Figura 4.2), a pesar de su margen de error del 39.78%.

- Resultados de las predicciones de CO

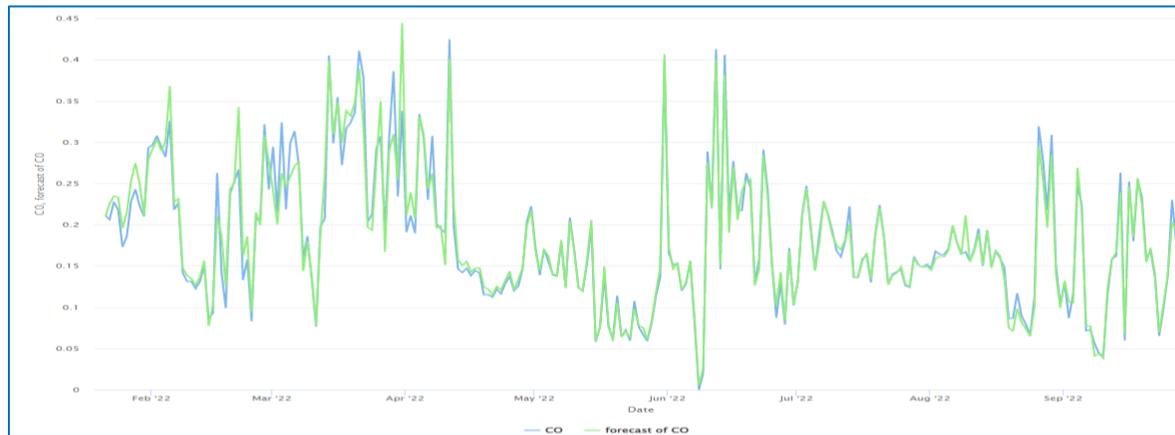


Figura 4.3 Comparación entre datos reales de CO y el método ARIMA

En la Figura 4.3, se puede ver que los datos actuales de CO, marcados de color azul en el gráfico y los datos del modelo ARIMA, marcados de color verde, cuentan con bastante similitud, lo que quiere decir que se pudo hacer una buena estimación de los datos a futuro (ver Figura 4.3), teniendo un margen de error del 5.85%.

- Resultados de las predicciones de CO_2



Figura 4.4 Comparación entre datos reales de CO_2 y el método ARIMA

En la Figura 4.4, se puede ver que los datos actuales de CO_2 , marcados de color azul en el gráfico y los datos predichos por el modelo ARIMA, marcados de color verde, cuentan con bastante similitud, lo que quiere decir que se pudo hacer una buena estimación de los datos a futuro (ver Figura 4.4), teniendo un margen de error del 13.63%.

4.5 Resultados del análisis de correlación entre variables

Con respecto al PM_{10} y al $PM_{2.5}$, sus concentraciones varían de acuerdo con su acumulación en la atmosfera.

En general, para las partículas de CO_2 y CO, se detectó un alto nivel de correlación con la temperatura del ambiente y la humedad, pues cuando la temperatura disminuía, se presentaban altos valores de dióxido de carbono, debido a que era tiempo de invierno y se estaban usando calefactores, pero cuando los niveles de temperatura crecían, los valores de monóxido de carbono aumentaban.

4.6 Resultados de las predicciones de árboles de decisión

En el estudio de cómo se clasifican los contaminantes en el aire interior, aplicando arboles de decisión potenciados por gradientes, se obtuvo la Tabla 4.2 con los resultados de los márgenes de error.

Tabla 4.2 Resultados de las predicciones futuras de los contaminantes del aire interior.

Predicción de:	Error cuadrático medio (RMSE)	Error absoluto medio (MAE)
$PM_{2.5}$ y PM_{10}	38.066	30.3%
CO	54.767	5.1%
CO_2	57.980	15.4%

Como se observa en la Tabla 4.2, al comparar los resultados de los árboles de decisión, para la clasificación de los contaminantes del aire interior, se puede notar que tienen margen de error mucho mayor que las predicciones hechas con el modelo ARIMA, aunque hay que recordar que los datos sobre los que se ha ajustado el modelo son los correspondientes a un solo año (ver tabla 4.2).

CAPÍTULO V. CONCLUSIONES

5.1 Conclusiones

A partir de los resultados obtenidos y partiendo de los objetivos planteados al inicio de este documento, se ha llegado a las siguientes conclusiones:

1. Fue posible obtener más conocimiento sobre las causas y los efectos de la contaminación del aire interior a través de la creación de este documento. Para lograr esto, se utilizaron las observaciones de las cuatro variables principales $PM_{2.5}$, PM_{10} , CO y CO_2 .
2. Durante el transcurso de este estudio, quedó claro que algunas cosas que parecían no tener nada que ver entre sí o que finalmente se creía que no tenían relación, de hecho, tenían una conexión. Por ejemplo, las variaciones de temperatura y humedad tuvieron un impacto tanto en los datos como en las relaciones entre los niveles de CO Y CO_2 .
3. Se concluye que el diseño del prototipo de red de contaminantes del aire interior permite probar la hipótesis de que es factible desarrollar una red de monitoreo de contaminantes mínimos que pueda identificar la presencia de gases tóxicos y material particulado (PM) suspendido en el aire al interior de ambientes cerrados. Este prototipo también permite la recopilación de datos meteorológicos para la predicción del nivel del nivel de exposición diaria a los contaminantes.
4. Los datos recopilados son suficientes para lograr el objetivo del proyecto, y se han establecido los pasos necesarios para combinar los datos originales de una manera que sea más adecuada para la investigación.
5. La biblioteca *pandas*, una de las herramientas gratuitas de análisis de datos más populares disponibles en la actualidad, se utilizó para casi todas las modificaciones realizadas a los datos y su visualización a lo largo de este proyecto. Pandas y RapidMiner se han utilizado a lo largo del proyecto para la interpretación, visualización y diseño de cada modelo ARIMA para el tamaño de los contaminantes del aire interior.
6. Al examinar las relaciones lineales entre las diversas magnitudes, se vieron claramente la utilidad y las implicaciones del coeficiente de correlación lineal.

7. También se ha llegado a comprender los mecanismos de RapidMiner para la estructura del código a la hora de diseñar el modelo ARIMA para las diferentes magnitudes de los contaminantes del aire interior.
8. Los modelos ARIMA han sido comparados entre ellos utilizando el error de sus predicciones con los datos reales y su proceso de aprendizaje para entender cómo se han llegado a esas conclusiones.
9. Si los datos muestran un patrón estacional, se utilizan las herramientas apropiadas para separarlos en los componentes estacionales (datos observados, tendencia, estacionalidad y residual) para estudiar la tendencia durante el período de tiempo especificado.
10. El proyecto ha conllevado una marcada mejora en la habilidad para utilizar las herramientas pertinentes y el lenguaje de programación Python, demostrando avances significativos.

5.2 Trabajos futuros

La investigación adicional que se requerirá para ampliar los hallazgos del estudio actual se puede dividir en las siguientes partes:

- Al utilizar las bibliotecas *Dash* y *Plotly* para la creación de gráficos interactivos, se puede ampliar la comprensión de Python mediante el desarrollo de un panel que se aloja en línea y permite a los usuarios consultar los resultados de este estudio.
- El uso de más módulos para expandir los datos en varios ambientes interiores donde es necesario clasificar y analizar la calidad del aire interior.
- Utilizar redes neuronales como modelos LSTM para comparar las predicciones realizadas por el modelo ARIMA.
- Establecer un criterio de parametrización y una arquitectura de red neuronal para mejorar la precisión de las predicciones.
- Conectar los prototipos a la red para que se envíen los datos en tiempo real a una base de datos dentro de la nube.
- Intercambiar la placa Arduino por Raspberry para tener mejor control de la información y se puedan visualizar los datos en tiempo real.

Referencias

- Villate Barrera, C. L. (5 de Mayo de 2017). *Repisitorio academico*. Obtenido de <https://repository.usta.edu.co/bitstream/handle/11634/10735/2018Villatecesar.pdf?sequence=1&isAllowed=y>
- Anturi Ramírez, H. (17 de Mayo de 2017). *El tiempo*. Obtenido de <https://www.eltiempo.com/colombia/medellin/jovenes-ganadores-de-roborave-international-2017-89770>
- Arduino. (14 de agosto de 2018). *Arduino*. Obtenido de <https://www.arduino.cc/>
- Ayala, N. (27 de Marzo de 2019). *sputniknews*. Obtenido de <https://mundo.sputniknews.com/ecologia/201903261086322742-arbol-robot-algas-purificacion-aire-mexico-biourban/>
- Becerra Gutiérrez, C. (18 de Julio de 2018). *Andina*. Obtenido de <https://andina.pe/agencia/noticia-peruanos-crean-robot-capaz-identificar-gases-toxicos-y-evitar-muerte-mineros-620234.aspx>
- Blanco Salazar, C. A. (23 de Febrero de 2018). *visor de documentos IMPI*. Recuperado el 10 de 04 de 2020, de <https://vidoc.impi.gob.mx/visor?usr=SIGA&texp=SI&tdoc=E&id=MX/a/2018/007659>
- Bombonabutano. (6 de Agosto de 2018). *bombonabutano.com*. Obtenido de <https://bombonabutano.com/info/gas-butano>
- Bosch. (20 de Enero de 2020). *bosch-sensortec*. Obtenido de <https://www.bosch-sensortec.com/products/environmental-sensors/gas-sensors-bme680/>
- Calvo, D. (8 de Diciembre de 2018). *diegocalvo*. Obtenido de <https://www.diegocalvo.es/perceptron-multicapa/>
- Castelazo, A. (7 de Junio de 2014). *codigoespagueti*. Obtenido de <https://codigoespagueti.com/noticias/perro-robot-capaz-fugas-de-gas/>
- Cheung, W.-F. (23 de Octubre de 2018). Un sistema de monitoreo de seguridad de construcción en tiempo real para gases peligrosos que integra redes de sensores inalámbricos y tecnologías de modelado de información de edificios. *MDPI*, 1-24.
- Crehana. (18 de Agosto de 2021). Crehana. *¿Qué es el análisis multivariado?*, 5.
- CRISTIAN, C. A. (6 de Junio de 2019). *Repositorio de la universidad piloto de colombia*. Obtenido de <http://repository.unipiloto.edu.co/bitstream/handle/20.500.12277/4880/Trabajo%20de%20grado5094.pdf?sequence=1>
- Datascientest. (13 de Diciembre de 2022). *Datascientest*. Obtenido de <https://datascientest.com/es/machine-learning-definicion-funcionamiento-usos>

El mundo lider mundial en español. (2 de Febrero de 2011). *El mundo.es*. Obtenido de https://www.elmundo.es/elmundo/2011/02/02/andalucia_malaga/1296641797.html

El patagonico. (05 de Octubre de 2011). Obtenido de <https://www.elpatagonico.com/fabricaron-un-robot-alertar-emergencias-los-yacimientos-n1406618>

Fernandez, M. (3 de Junio de 2019). *El español*. Obtenido de https://www.elespanol.com/omicron/tecnologia/20190603/robot-medioambiental-ahorra-energia-imitando-perezosos/403460855_0.html

Gavilánez, C. S. (5 de Febrero de 2021). *Repositorio Universidad Técnica de Ambato*. Obtenido de <https://repositorio.uta.edu.ec/handle/123456789/32088?mode=full>

GEOSLAM. (18 de Enero de 2019). *Geoslam*. Obtenido de <https://geoslam.com/blog/2019/01/18/what-is-slam/>

Gupta, T. (4 de Enero de 2017). *Towards data science*. Obtenido de <https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7>

IARR. (5 de Julio de 2020). *IARRBOOK*. Obtenido de <https://iaarbook.github.io/vision-por-computadora/>

IBM knowledge center. (16 de Mayo de 2019). *ibm.com*. Obtenido de https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html

IMECA. (21 de Octubre de 2018). *Gob.mx*. Obtenido de <https://www.gob.mx/comisionambiental/articulos/imeca-indice-metropolitano-de-la-calidad-del-aire?idiom=es>

Instituto Nacional de Ecología y Cambio climático. (17 de Agosto de 2022). *sinaica*. Obtenido de <https://sinaica.inecc.gob.mx/>

Instituto para la salud geoambiental. (15 de Mayo de 2020). *saludgeoambiental.org/*. Obtenido de <https://www.saludgeoambiental.org/material-particulado>

Kenton. (30 de Noviembre de 2020). *Investopedia*. Obtenido de <https://www.investopedia.com/terms/s/seasonality.asp>

Lis data solutions. (15 de Febrero de 2023). *Datasolutions*. Obtenido de <https://www.lisdatasolutions.com/es/que-es-rapidminer/>

Maria Puig-Arnavat, J. C. (12 de Agosto de 2015). *sciencedirect*. Obtenido de <https://www.sciencedirect.com/topics/engineering/backpropagation-algorithm>

Marso, S. (6 de Abril de 2020). Predicting financial distress using hybrid feedforward neural network. *ELSEVIER*, 2-7.

Martinez, A. (26 de Agosto del 2014). Los gases industriales y su clasificación. *Desenchufados*, 1-2.

Master magazine. (12 de Febrero del 2018). Deficion de datos. *Sistemas*, 1-2.

- Menéndez, A. d. (15 de Junio de 2015). *http://arantxa.ii.uam.es/*. Obtenido de <http://arantxa.ii.uam.es/~jms/pfcsteleco/lecturas/20150625AlmudenaDePazMenendez.pdf>
- Microsoft. (1 de Enero de 2019). *docs.microsoft.com*. Obtenido de <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>
- Microsoft. (15 de Septiembre de 2022). *Microsoft*. Obtenido de <https://learn.microsoft.com/es-es/microsoft-365/education/data-streamer/>
- Minitab. (14 de Abril de 2022). *support minitab*. Obtenido de <https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/interpret-the-results/key-results/#:~:text=La%20tabla%20prueba%20aumentada%20de,de%20evaluar%20la%20hip%C3%B3tesis%20nula>.
- Noticias de ciencia y tecnología*. (11 de Octubre de 2011). Obtenido de <https://noticiadelaciencia.com/art/2413/robot-para-tareas-de-rescate-de-mineros>
- Núñez Caraballo, V. (12 de Enero de 2014). Metodología de diagnóstico técnico ambiental en fuentes fijas industriales para prevenir y mitigar los Episodios Críticos de Contaminación Atmosférica y el impacto a la salud humana. Santa clara: UNIVERSIDAD CENTRAL MARTA ABREU DE LAS VILLAS.
- Observatorio de salud y medio ambiente de andalucía. (15 de Octubre del 2011). Calidad del aire interior. *Junta de Andalucía*, 11-15.
- ORACLE. (16 de Mayo de 2020). *oracle mexico*. Obtenido de <https://www.oracle.com/mx/artificial-intelligence/what-is-artificial-intelligence.html>
- Organización mundial de las Naciones Unidas para la Alimentación y Agricultura. (10 de Agosto del 2019). La contaminación del suelo una realidad. *Alianza mundial del suelo*, 20-27.
- Organización Panamericana de la Salud. (18 de Enero de 2018). *OPS*. Obtenido de https://www.paho.org/hq/index.php?option=com_content&view=article&id=14454:ambient-and-household-air-pollution-and-health-frequently-asked-questions&Itemid=72243&lang=es
- Ortega, L. P. (10 de Junio de 2015). *CTMA*. Obtenido de Luis P. Ortega
- Pastor, R. (14 de Septiembre de 2018). *styd.net*. Obtenido de <https://styd.net/que-es-y-para-que-sirve-sql/>
- RadioShuttle. (7 de Julio de 2020). *RadioShuttle*. Obtenido de <https://www.radioshuttle.de/es/radioshuttle-es/ejemplos/sensor-ambiental/>
- Raffino, M. E. (10 de Junio de 2020). *Concepto.de*. Obtenido de <https://concepto.de/contaminacion-sonora/>
- Raspberry Pi. (12 de Mayo de 2020). *raspberrypi.org*. Obtenido de <https://www.raspberrypi.org/help/what-%20is-a-raspberry-pi/>

- Revista Universidad y Sociedad. (8 de Abril del 2020). Monitorización de gases contaminantes en ambientes cerrados usando WSN para la toma de acciones preventivas. *Revista Universidad y Sociedad*, 1-9.
- Rouse, M. (10 de Julio de 2018). *techtarget*. Obtenido de <https://searchdatacenter.techtarget.com/es/definicion/Base-de-datos>
- Sanchez, J. (30 de Enero de 2020). *Ecología verde*. Recuperado el 10 de 04 de 2020, de <https://www.ecologiaverde.com/que-es-la-contaminacion-ambiental-y-sus-tipos-1336.html>
- Soloaga, A. (19 de Octubre de 2018). Obtenido de akademus: <https://www.akademus.es/blog/programacion/principales-usos-python/>
- sostenibilidadedp. (10 de Octubre de 2016). *www.sostenibilidadedp.es*. Obtenido de <https://www.sostenibilidadedp.es/pages/index/que-es-el-ruido>
- Soto, A. A. (20 de Junio del 2019). Monitoreo de contaminantes en el aire y sus efectos. *REVISTA UNIVERSITARIA*.
- Synergy codes. (1 de Marzo de 2023). *Synergy codes*. Obtenido de <https://synergycodes.com/glossary/what-is-density-plot/>
- Ucha, F. (12 de Noviembre de 2019). *Definición ABC*. Obtenido de <https://www.definicionabc.com/general/aire.php>
- Universidad de arizona. (15 de Marzo de 2020). *aska*. Obtenido de <https://askabiologist.asu.edu/anatomia-de-la-neurona>