

Pontificia Universidad Católica del Ecuador

Facultad de Ingeniería

Maestría en Ciencia de Datos

Desarrollo de un modelo que prediga el nivel de contaminación del aire en la
ciudad de Quito.

Mora González Bairon Hernán

2023

Índice de contenidos

| | |
|--|----|
| 1. Introducción | 7 |
| 1.1. Problemática..... | 8 |
| 1.2. Objetivos | 9 |
| 1.2.1. Objetivo general: | 9 |
| 1.2.2. Objetivos específicos:..... | 10 |
| 1.3. Justificación..... | 10 |
| 2. Marco teórico..... | 12 |
| 2.1. Contaminación atmosférica..... | 12 |
| 2.2 Tipos de contaminantes..... | 13 |
| 2.2.1. Monóxido de carbono (CO)..... | 14 |
| 2.2.2. Dióxido de Nitrógeno (NO ₂)..... | 16 |
| 2.2.3. Ozono (O ₃) | 17 |
| 2.2.4. Partículas menores a 2.5 micras (PM _{2.5})..... | 19 |
| 2.2.5. Partículas menotes a 10 micras (PM ₁₀) | 20 |
| 2.2.6. Dióxido de azufre (SO ₂)..... | 22 |
| 2.2.7. Dirección del Viento (DIR)..... | 23 |
| 2.2.8. Datos Humedad relativa (HUM) | 23 |
| 2.2.9. Datos radiación ultravioleta (IUV)..... | 24 |
| 2.2.10. Datos precipitación (LLU)..... | 26 |
| 2.2.11. Presión barométrica (PRE) | 27 |
| 2.2.12. Temperatura media (TMP)..... | 28 |
| 2.2.13. Velocidad del viento (VEL) | 29 |
| 2.3 Sensores de medición..... | 30 |
| 2.4 Machine Learning | 30 |

| | |
|--|----|
| 2.4.1 Descripción e importancia de machine learning..... | 32 |
| 2.4.2 Describir el Pipeline de un proceso de machine learning..... | 32 |
| 2.4.3 Descripción de Deep learning..... | 33 |
| 2.5 Estudios previos | 34 |
| 3. Metodología | 36 |
| 3.1. Tipo de estudio | 36 |
| 3.2. Descripción de datos y fuente de información..... | 36 |
| 3.3. Descripción de modelos de machine learning supervisado..... | 38 |
| 3.4. Descripción de redes neuronales | 40 |
| 3.5. Métricas de evaluación del modelo | 40 |
| 4. Análisis de resultados | 41 |
| Bibliografía | 72 |

Índice de figuras

| | |
|--|----|
| Figura 1. Características de la maquina en Google Colab | 41 |
| Figura 2. Repositorio y distribución de datos disponibles | 42 |
| Figura 3. Descarga de datos de contaminación..... | 43 |
| Figura 4. Descompresión de datos en la carpeta XLSX | 43 |
| Figura 5. Importación individual de datos | 44 |
| Figura 6. Renombrado de campos | 44 |
| Figura 7. Unión de datos | 44 |
| Figura 8. Eliminación de valores perdidos | 45 |
| Figura 9. Partes por millón de partículas de 2.5 micras | 46 |
| Figura 10. Microgramos de ozono por metro cuadrado..... | 46 |
| Figura 11. Microgramos de dióxido de azufre por metro cúbico | 47 |
| Figura 12. Microgramos de dióxido de nitrógeno | 48 |
| Figura 13. Miligramos de monóxido de carbono por metro cúbico | 48 |
| Figura 14. Descriptivos de las características que determinan la calidad del aire | 49 |
| Figura 15. Partes por millón de partículas de 2.5 micras por zona..... | 50 |
| Figura 16. Concentración de dióxido de azufre por zona | 50 |
| Figura 17. Concentración de dióxido de nitrógeno por zona..... | 51 |
| Figura 18. Concentración de ozono por zona | 52 |
| Figura 19. Concentración de monóxido de carbono por zona | 52 |
| Figura 20. Matriz de correlación..... | 54 |
| Figura 21. Matriz de dispersión de compuestos contaminantes | 55 |
| Figura 22. Gráfico de caja de las partes por millón de partículas de 2.5 micras por zona | 56 |
| Figura 23. Gráfico de la concentración de ozono por zona | 56 |
| Figura 24. Concentración de dióxido de nitrógeno por zona..... | 57 |
| Figura 25. Concentración del dióxido de azufre por zona | 58 |
| Figura 26. Concentración de monóxido de carbono por zona | 58 |
| Figura 27. Topología de la red neuronal..... | 60 |
| Figura 28. Valores predichos VS valores reales – Modelo lineal..... | 60 |
| Figura 29. Valores predichos VS valores reales – Modelo SVR..... | 61 |
| Figura 30. Valores predichos VS valores reales – Modelo de árbol de predicción | 62 |

| | |
|---|----|
| Figura 31. Valores predichos VS valores reales – Modelo KNN | 62 |
| Figura 32. Valores predichos VS valores reales – Modelo de red neuronal | 63 |

Índice de tablas

| | |
|---|----|
| Tabla 1. Configuración de hiperparámetros..... | 59 |
| Tabla 2. Métricas de evaluación de los modelos de predicción..... | 64 |

1. Introducción

La calidad del aire es un tema de creciente preocupación en las áreas urbanas debido a sus impactos negativos en la salud humana y el medio ambiente. Las ciudades experimentan altos niveles de contaminación atmosférica debido a la concentración de actividades industriales, el aumento del tráfico vehicular y otros factores antropogénicos. Quito, la capital de Ecuador, no es una excepción a esta problemática.

El monitoreo y la predicción de la contaminación del aire en Quito son fundamentales para comprender la naturaleza y los patrones de la calidad del aire en la ciudad, así como para implementar medidas adecuadas de mitigación y control. Para lograrlo, es necesario contar con un modelo preciso que pueda predecir el nivel de contaminación del aire en función de los datos históricos disponibles.

En este estudio, se propone el desarrollo de un modelo predictivo basado en estadísticas de bases de datos de años anteriores para predecir el nivel de contaminación del aire en la ciudad de Quito. El enfoque meso implica considerar la ciudad en su conjunto, abordando la variabilidad espacial y temporal de los contaminantes atmosféricos en diferentes zonas de la ciudad y en diferentes momentos del día.

El modelo meso utilizará una combinación de datos meteorológicos, datos de emisiones y datos de calidad del aire recopilados a lo largo de varios años. Estos datos incluirán variables como la temperatura, la humedad, la velocidad del viento, la precipitación, las emisiones industriales y vehiculares, así como las concentraciones históricas de contaminantes como partículas PM2.5, dióxido de nitrógeno (NO₂), ozono (O₃) y otros contaminantes relevantes.

Además del enfoque meso, es importante considerar el ámbito micro en el desarrollo de un modelo predictivo para la contaminación del aire en la ciudad de Quito. El ámbito micro se refiere a la escala más localizada, donde se tienen en cuenta factores como la topografía, la densidad de población y las características específicas de los diferentes barrios y áreas urbanas.

La contaminación del aire puede variar significativamente dentro de una ciudad debido a la presencia de fuentes de emisión específicas, como fábricas, áreas industriales, zonas de alta densidad de tráfico o áreas verdes. Estas variaciones pueden influir en la calidad del aire en diferentes vecindarios y en diferentes momentos del día.

El objeto de estudio de esta investigación científica es el nivel de contaminación del aire en la ciudad de Quito. El objetivo es comprender la dinámica de la contaminación atmosférica en la ciudad, identificar los factores que influyen en ella y desarrollar un modelo predictivo que pueda estimar con precisión los niveles de contaminación en función de datos estadísticos disponibles recopilados de años anteriores.

El campo de estudio abarca la calidad del aire en Quito, centrándose en la recopilación, análisis e interpretación de datos relacionados con la contaminación atmosférica. Se explorarán diferentes aspectos dentro de este campo, como la variabilidad espacial y temporal de los contaminantes, las fuentes de emisión, los factores meteorológicos y las características urbanas que influyen en la calidad del aire. Este modelo utilizará técnicas de análisis de datos y aprendizaje automático para establecer relaciones entre los factores que influyen en la contaminación del aire y los niveles de contaminantes registrados en el pasado.

1.1. Problemática

La problemática principal de esta investigación es la alta contaminación del aire en la ciudad de Quito, lo cual genera diversos efectos negativos en el entorno afectado. Estos efectos incluyen:

1. **Impacto en la salud pública:** La contaminación del aire en Quito provoca un aumento en los índices de enfermedades respiratorias, cardiovasculares y otras afecciones relacionadas. La población expuesta a altos niveles de contaminantes sufre de problemas respiratorios, irritación de los ojos, alergias y un mayor riesgo de desarrollar enfermedades crónicas.
2. **Degradación del medio ambiente:** Los contaminantes atmosféricos afectan negativamente los ecosistemas locales, incluyendo los bosques y las áreas naturales cercanas a la ciudad. La vegetación y la fauna se ven perjudicadas, lo que reduce la biodiversidad y altera los ciclos naturales.
3. **Cambio climático:** La contaminación del aire contribuye al calentamiento global y al cambio climático. La emisión de gases de efecto invernadero, como el dióxido de carbono (CO₂), agravada por la mala calidad del aire, afecta el equilibrio del clima local y global, con impactos en el clima, la agricultura y la disponibilidad de recursos naturales.

Problemas secundarios:

1. Fuentes de emisión de contaminantes: Se identifica una falta de control y regulación efectiva de las fuentes de emisión de contaminantes en la ciudad de Quito. Esto incluye emisiones de industrias, fábricas, vehículos de transporte, quema de residuos.
2. Planificación urbana inadecuada: La falta de una planificación urbana integral y sostenible en Quito contribuye a la concentración de fuentes de contaminación, como zonas industriales y vías de tráfico congestionadas. La distribución inadecuada de áreas verdes y espacios abiertos también puede afectar la calidad del aire.
3. Falta de conciencia y educación ambiental: Existe una falta de conciencia y educación ambiental en la población de Quito en relación con los efectos negativos de la contaminación del aire y las medidas individuales y colectivas que se pueden tomar para mitigarla. Se presenta a continuación un diagrama de espina de pescado para visualizar las causas y efectos de la problemática principal y los problemas secundarios identificados:

Problema principal: Alta contaminación del aire en Quito.

Efectos:

- Impacto en la salud pública.
- Degradación del medio ambiente.
- Cambio climático.

Causas:

- Fuentes de emisión de contaminantes.
- Planificación urbana inadecuada.
- Falta de conciencia y educación ambiental.

1.2. Objetivos

1.2.1. Objetivo general:

El objetivo general de esta investigación es desarrollar un modelo predictivo que pueda estimar con precisión el nivel de contaminación del aire en la ciudad de Quito, con el fin de contribuir a la mitigación de la problemática de alta contaminación atmosférica.

1.2.2. Objetivos específicos:

1. Identificar y caracterizar las fuentes de emisión de contaminantes en la ciudad de Quito, incluyendo industrias, vehículos de transporte.
2. Evaluar el impacto de la planificación urbana en la concentración de contaminantes y en la calidad del aire en diferentes áreas de la ciudad.
3. Recopilar y analizar datos estadísticos de bases de datos disponibles de años anteriores relacionados con la calidad del aire, las condiciones meteorológicas, las emisiones y otros factores relevantes en Quito.
4. Desarrollar un modelo predictivo utilizando técnicas de análisis de datos y aprendizaje automático, que permita estimar con precisión el nivel de contaminación del aire en diferentes zonas de la ciudad, considerando factores meso y micro.
5. Validar y evaluar el modelo predictivo utilizando datos actuales de calidad del aire y comparando las predicciones del modelo con las mediciones reales.
6. Proponer medidas de mitigación y control basadas en los resultados obtenidos, con el objetivo de reducir la contaminación del aire en Quito y mejorar la calidad del aire en beneficio de la salud pública y el medio ambiente.

1.3. Justificación

Es importante investigar sobre el tema de la contaminación del aire en la ciudad de Quito debido a las siguientes razones:

1. Salud pública: La contaminación del aire tiene un impacto directo en la salud de la población. Al investigar y abordar este problema, se busca reducir los riesgos de enfermedades respiratorias y cardiovasculares, mejorando la calidad de vida de los habitantes de la ciudad.
2. Medio ambiente y sostenibilidad: La contaminación del aire no solo afecta a las personas, sino también al medio ambiente. Al investigar y mitigar este problema, se contribuye a la conservación de los ecosistemas locales, incluyendo bosques, ríos y áreas naturales cercanas. Esto favorece la biodiversidad, la provisión de servicios eco sistémicos y la preservación de los recursos naturales, siendo fundamental para la sostenibilidad ambiental a largo plazo.

3. Cumplimiento de regulaciones y normativas: Investigar y solucionar el problema de la contaminación del aire en Quito permite cumplir con las regulaciones y normativas nacionales e internacionales relacionadas con la calidad del aire. Esto evita sanciones y multas, además de mejorar la reputación y el prestigio institucional tanto a nivel local como global.

4. Desarrollo económico y competitividad: Al mejorar la calidad del aire, se fomenta el desarrollo económico sostenible de la ciudad. Una buena calidad del aire puede atraer inversiones, promover el turismo y favorecer la instalación de nuevas empresas.

La investigación y posible solución del problema de la contaminación del aire en Quito se realizará con el objetivo de proteger la salud de la población, cumplir con las regulaciones ambientales, asumir una responsabilidad social y ambiental, mejorar la calidad de vida de la comunidad y contribuir al desarrollo sostenible de la ciudad.

2. Marco teórico

2.1. Contaminación atmosférica

La contaminación atmosférica se refiere a la presencia de sustancias o partículas nocivas en la atmósfera terrestre que pueden tener efectos adversos en la salud humana, en la vida silvestre y en el medio ambiente en general, aquí hay algunos puntos clave sobre la contaminación atmosférica:

- Fuentes de contaminación atmosférica: La contaminación atmosférica puede provenir de diversas fuentes, incluyendo la quema de combustibles fósiles (como el petróleo y el carbón) en industrias, automóviles y centrales eléctricas, la agricultura intensiva, la deforestación, las erupciones volcánicas y otros procesos naturales (Narváez, 2021).
- Contaminantes atmosféricos: Los contaminantes comunes incluyen dióxido de azufre (SO₂), dióxido de nitrógeno (NO₂), monóxido de carbono (CO), partículas suspendidas en el aire (PM_{2.5} y PM₁₀), ozono troposférico (O₃), compuestos orgánicos volátiles (COV) y plomo, entre otros (Paredes, 2020).
- Efectos en la salud humana: La exposición a la contaminación atmosférica puede causar problemas de salud, como enfermedades respiratorias (asma, bronquitis, enfisema), enfermedades cardiovasculares, cáncer de pulmón y problemas neurológicos. Además, puede tener efectos perjudiciales en el desarrollo cognitivo en niños (Aránguez, 2019).
- Efectos en el medio ambiente: La contaminación del aire puede dañar los ecosistemas, afectar a la biodiversidad, acidificar cuerpos de agua y suelos, y contribuir al cambio climático al aumentar las concentraciones de gases de efecto invernadero (Pallo, 2021).
- Regulación y mitigación: Muchos países han implementado regulaciones para controlar la contaminación del aire, estableciendo límites máximos de emisiones para las industrias y los vehículos. Además, se promueve el uso de tecnologías más limpias y renovables, como energía solar y eólica, para reducir la dependencia de combustibles fósiles (Placeres, 2020).
- Participación ciudadana: La conciencia pública y la participación son fundamentales para abordar la contaminación atmosférica. La reducción del uso de vehículos personales, la promoción del transporte público, el reciclaje y la adopción de prácticas más sostenibles pueden ayudar a mitigar la contaminación del aire.

- Monitoreo de la calidad del aire: Se realizan mediciones constantes de la calidad del aire en todo el mundo para evaluar los niveles de contaminantes y tomar medidas adecuadas cuando se superan los límites permitidos.

La contaminación atmosférica es un problema grave que requiere esfuerzos a nivel global para abordar sus causas y mitigar sus efectos. La conciencia pública y la acción colectiva desempeñan un papel crucial en la protección de la salud humana y el medio ambiente contra la contaminación del aire.

2.2 Tipos de contaminantes

Los contaminantes atmosféricos pueden clasificarse en varios tipos según sus características y fuentes de emisión. Aquí tienes una descripción de los tipos principales de contaminantes atmosféricos:

Contaminantes primarios: Estos contaminantes son emitidos directamente por fuentes específicas y pueden estar en forma gaseosa o de partículas. Algunos ejemplos de contaminantes primarios incluyen (Sango, 2018):

- Dióxido de azufre (SO₂): Emitido por la quema de combustibles fósiles, como el carbón y el petróleo, y puede contribuir a la formación de lluvia ácida.
- Dióxido de nitrógeno (NO₂): Producido principalmente por la quema de combustibles en vehículos y plantas de energía, y puede contribuir a la formación de ozono troposférico y la lluvia ácida.
- Monóxido de carbono (CO): Emitido por vehículos y la quema de biomasa, es un gas tóxico que puede afectar la salud humana.
- Partículas suspendidas en el aire (PM_{2.5} y PM₁₀): Partículas sólidas o líquidas en suspensión en el aire que pueden ser emitidas por la quema de combustibles, procesos industriales y actividades de construcción. Son perjudiciales para la salud respiratoria.
- Compuestos orgánicos volátiles (COV): Son emitidos por diversas fuentes, como vehículos, industrias y productos químicos. Contribuyen a la formación de ozono troposférico y pueden ser tóxicos.

Contaminantes secundarios: Estos contaminantes se forman en la atmósfera a partir de reacciones químicas entre contaminantes primarios y otros compuestos. Algunos ejemplos incluyen (Malagón, 2011):

- Ozono troposférico (O₃): Se forma cuando los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV) reaccionan en presencia de luz solar. A nivel del suelo, el ozono puede ser perjudicial para la salud.
- Partículas secundarias: Las partículas pueden formarse en la atmósfera a partir de precursores gaseosos como el dióxido de azufre y los compuestos orgánicos volátiles.

Contaminantes orgánicos persistentes (COP): Estos son compuestos químicos orgánicos que pueden persistir en el ambiente durante largos períodos y pueden acumularse en los tejidos vivos. Ejemplos de COP incluyen los bifenilos policlorados (PCB) y los pesticidas organoclorados, como el DDT (Chiluisa, 2023).

- Metales pesados: Estos contaminantes incluyen metales como el plomo, el mercurio, el cadmio y el arsénico, que pueden ser liberados por actividades industriales y mineras. Pueden ser tóxicos para los seres humanos y los ecosistemas.
- Contaminantes radiactivos: Incluyen sustancias radiactivas como el radón y los productos de desintegración nuclear, que pueden ser emitidos por actividades nucleares y pueden ser peligrosos para la salud.

Cada tipo de contaminante tiene sus propias fuentes de emisión, efectos en la salud y el medio ambiente, y métodos de control específicos. La gestión y reducción de estos contaminantes son esenciales para preservar la calidad del aire y mitigar los impactos negativos en la salud humana y el ecosistema.

2.2.1. Monóxido de carbono (CO)

El monóxido de carbono (CO) es un gas incoloro, inodoro y altamente tóxico que se encuentra en la atmósfera. A pesar de que no tiene un olor característico, es peligroso porque puede ser letal para los seres humanos y otros organismos cuando se inhala en concentraciones suficientemente altas (Servicio de Salud Pública, 2012).

Fuentes de emisión:

- **Combustión incompleta:** El CO se produce principalmente durante la combustión incompleta de materiales que contienen carbono. Esto incluye la quema de combustibles fósiles como el gas natural, la gasolina, el petróleo y el carbón en vehículos de motor, calderas, estufas, chimeneas y otras aplicaciones industriales y domésticas.
- **Incendios:** Los incendios en edificios o en la naturaleza también pueden ser una fuente significativa de monóxido de carbono, liberándolo en el aire.

Riesgos para la salud:

- **Toxicidad:** El CO se une a la hemoglobina en los glóbulos rojos de la sangre con mayor afinidad que el oxígeno, lo que reduce la capacidad de la sangre para transportar oxígeno a los tejidos del cuerpo. La exposición al CO puede causar síntomas como dolor de cabeza, mareos, náuseas, confusión y debilidad. En concentraciones elevadas o exposiciones prolongadas, puede ser mortal (Aránguez, 2019).
- **Efectos a largo plazo:** La exposición crónica al CO, incluso a niveles bajos, puede tener efectos negativos en la salud, como problemas neurológicos, daño cardiovascular y efectos perjudiciales en el desarrollo fetal.

Prevención y control:

- **Detectores de CO:** Se recomienda encarecidamente la instalación de detectores de monóxido de carbono en el hogar, especialmente cerca de las áreas de dormitorio. Estos dispositivos pueden alertar a las personas cuando los niveles de CO en el aire alcanzan niveles peligrosos (Aránguez, 2019).
- **Mantenimiento adecuado:** La prevención de la intoxicación por CO implica mantener en buen estado los sistemas de calefacción, ventilación y aire acondicionado, así como los electrodomésticos de gas y leña. Además, es importante evitar el uso de parrillas de carbón en espacios cerrados y mantener áreas bien ventiladas cuando se usen generadores portátiles u otros dispositivos que generen CO.
- **Educación sobre la seguridad:** La conciencia pública y la educación son clave para prevenir la exposición al CO. Las personas deben conocer los riesgos asociados con la combustión incompleta y tomar medidas para protegerse a sí mismas y a sus familias.

El monóxido de carbono es una amenaza seria para la salud que puede evitarse con medidas adecuadas de prevención y control. La detección temprana y la atención médica son esenciales en caso de exposición a niveles peligrosos de CO.

2.2.2. Dióxido de Nitrógeno (NO₂)

El dióxido de nitrógeno (NO₂) es un gas tóxico que forma parte de la familia de los óxidos de nitrógeno (NO_x). Se encuentra en la atmósfera como resultado de diversas actividades humanas y procesos naturales. Aquí tienes información importante sobre el dióxido de nitrógeno (NO₂) (Gallegos, 2018):

Fuentes de emisión:

- **Combustión de combustibles fósiles:** El NO₂ se emite principalmente a través de la combustión de combustibles fósiles, como gasolina y diésel en vehículos automotores, y la quema de carbón y petróleo en centrales eléctricas e instalaciones industriales (Organización Mundial de la Salud, 2019).
- **Procesos industriales:** Algunas industrias, como las plantas de producción de acero y químicas, pueden liberar NO₂ como subproducto de sus procesos.
- **Reacciones atmosféricas:** El NO₂ también se forma en la atmósfera como resultado de reacciones químicas entre el óxido nítrico (NO) y el oxígeno atmosférico en presencia de luz solar.

Riesgos para la salud:

- **Irritación respiratoria:** La exposición a concentraciones elevadas de NO₂ puede causar irritación en los pulmones y las vías respiratorias, lo que resulta en síntomas como tos, dificultad para respirar y agravamiento de enfermedades respiratorias preexistentes, como el asma y la bronquitis (Ibarlucía, 2019).
- **Impacto en la salud cardiovascular:** Se ha demostrado que la exposición crónica a niveles moderados de NO₂ está relacionada con un mayor riesgo de enfermedades cardiovasculares, incluidos problemas como ataques cardíacos y enfermedades del corazón.

- Efectos en la salud infantil: Los niños son particularmente vulnerables a los efectos del NO₂, ya que pueden experimentar un desarrollo pulmonar deficiente y un mayor riesgo de infecciones respiratorias.
- Contribución a la formación de ozono troposférico: El NO₂ es un precursor del ozono troposférico (O₃), un contaminante atmosférico secundario. En presencia de luz solar, el NO₂ puede reaccionar con los compuestos orgánicos volátiles (COV) para formar ozono a nivel del suelo, lo que puede aumentar la contaminación del aire y empeorar los problemas de salud respiratoria.

Control y mitigación:

- Regulación de emisiones: Muchos países han implementado regulaciones estrictas para controlar las emisiones de NO₂ de vehículos y fuentes industriales. Estas regulaciones incluyen estándares de emisiones más bajos y tecnologías de control de la contaminación (Organización Mundial de la Salud, 2019).
- Promoción del transporte limpio: Fomentar el uso de vehículos eléctricos, híbridos y de bajas emisiones, así como mejorar el transporte público y las infraestructuras para bicicletas y peatones, puede ayudar a reducir las emisiones de NO₂ en áreas urbanas.
- Transición a energías limpias: La transición de la energía sucia a fuentes de energía más limpias y renovables puede reducir las emisiones de NO₂ asociadas con la generación de energía.

La reducción de las emisiones de dióxido de nitrógeno es fundamental para mejorar la calidad del aire y proteger la salud humana y el medio ambiente. La combinación de regulaciones, tecnologías más limpias y cambios en la movilidad y la generación de energía son pasos importantes para lograr este objetivo.

2.2.3. Ozono (O₃)

El ozono (O₃) es una forma de oxígeno que consiste en tres átomos de oxígeno en lugar de los dos que se encuentran en el oxígeno molecular normal (O₂). El ozono se encuentra tanto en la alta atmósfera como en la troposfera, la capa más baja de la atmósfera terrestre. Tiene roles y características diferentes en estas dos regiones (Comunidad de Madrid,, 2021):

Ozono estratosférico (Ozono Bueno):

La mayor parte del ozono en la atmósfera se encuentra en la estratosfera, en una región conocida como la capa de ozono. Esta capa se encuentra a altitudes de aproximadamente 10 a 50 kilómetros sobre la Tierra.

El ozono estratosférico es esencial para la vida en la Tierra, ya que actúa como un escudo que absorbe y bloquea la radiación ultravioleta (UV) del sol. Esta radiación UV es perjudicial para los seres vivos y puede causar daño en la piel, cataratas y cáncer de piel (Chicón, 2023).

A lo largo de las décadas, ha habido una preocupación creciente sobre la destrucción de la capa de ozono debido a la liberación de sustancias químicas llamadas clorofluorocarbonos (CFCs) y otros compuestos que degradan el ozono estratosférico. Esto ha llevado a la adopción del Protocolo de Montreal y acuerdos internacionales para eliminar gradualmente la producción y el uso de estas sustancias.

Ozono Troposférico (Ozono Malo):

El ozono troposférico se encuentra a nivel del suelo o en las capas más bajas de la atmósfera, en áreas densamente pobladas y cerca de las fuentes de contaminación.

A diferencia del ozono estratosférico, el ozono troposférico no es beneficioso para la salud humana. Se forma a partir de reacciones químicas complejas entre precursores como el dióxido de nitrógeno (NO₂) y los compuestos orgánicos volátiles (COV) en presencia de luz solar (Aránguez, 2019).

El ozono troposférico es un contaminante atmosférico secundario y es un componente importante de la contaminación del aire en áreas urbanas. Puede causar problemas respiratorios, irritación de los ojos y agravar condiciones de salud preexistentes como el asma.

La formación de ozono troposférico se ve agravada por la presencia de altas concentraciones de NO₂ y COV, que a menudo provienen de fuentes industriales, vehículos de motor y procesos de combustión.

En resumen, mientras que el ozono estratosférico es beneficioso ya que protege contra la radiación ultravioleta dañina, el ozono troposférico es un contaminante atmosférico que puede ser perjudicial para la salud humana. El control de las emisiones de precursores, como NO₂ y COV, es esencial

para reducir la formación de ozono troposférico y mejorar la calidad del aire en áreas urbanas (Gallegos, 2018).

2.2.4. Partículas menores a 2.5 micras (PM2.5)

Las partículas menores a 2.5 micras, abreviadas como PM2.5, son partículas sólidas o líquidas suspendidas en el aire que tienen un diámetro aerodinámico igual o menor a 2.5 micrómetros (μm). Son un tipo de contaminante atmosférico particulado y se consideran partículas finas debido a su tamaño diminuto. Estas partículas son lo suficientemente pequeñas como para permanecer suspendidas en el aire durante largos períodos de tiempo y penetrar profundamente en los pulmones cuando se inhalan (Sanitaria, 2019).

Fuentes de emisión:

Las PM2.5 pueden tener diversas fuentes, tanto naturales como antropogénicas (causadas por actividades humanas), que incluyen (Chicón, 2023):

- **Combustión de combustibles Fósiles:** La quema de carbón, petróleo y gas natural en centrales eléctricas, vehículos de motor y calentadores es una de las principales fuentes de PM2.5.
- **Emisiones industriales:** Las industrias que liberan partículas finas incluyen plantas manufactureras, fundiciones y refinerías.
- **Actividades de construcción:** Las operaciones de construcción, como el movimiento de tierras y la demolición, pueden generar partículas finas en el aire.
- **Emisiones de agricultura:** La agricultura, a través de prácticas como la quema de rastrojos y la aplicación de fertilizantes, puede contribuir a la liberación de PM2.5.
- **Polvo y partículas naturales:** Parte de las PM2.5 en la atmósfera proviene de fuentes naturales, como polvo del suelo, erupciones volcánicas e incendios forestales.

Impactos en la salud:

La exposición a concentraciones elevadas de PM2.5 puede tener graves consecuencias para la salud humana, incluyendo:

- Problemas respiratorios: Las PM2.5 pueden penetrar profundamente en los pulmones, causando irritación de las vías respiratorias, exacerbando el asma y aumentando el riesgo de infecciones respiratorias.
- Problemas cardiovasculares: La exposición crónica a PM2.5 está relacionada con enfermedades cardiovasculares, incluyendo ataques cardíacos, arritmias y enfermedades del corazón.
- Cáncer de pulmón: Se ha vinculado la exposición a largo plazo a PM2.5 con un mayor riesgo de cáncer de pulmón.
- Mortalidad prematura: La contaminación del aire, en gran parte debido a PM2.5, está asociada con la mortalidad prematura en poblaciones expuestas.

Regulación y monitoreo:

Muchos países tienen regulaciones y estándares de calidad del aire para limitar las concentraciones de PM2.5 en el ambiente. El monitoreo constante de la calidad del aire se lleva a cabo para evaluar los niveles de PM2.5 y tomar medidas adecuadas cuando se superan los límites permitidos.

La reducción de las emisiones de PM2.5 y la adopción de tecnologías más limpias en sectores como la industria y el transporte son esenciales para mitigar los efectos negativos en la salud humana y el medio ambiente asociados con estas partículas finas.

2.2.5. Partículas menores a 10 micras (PM10)

Las partículas menores a 10 micras, abreviadas como PM10, son partículas sólidas o líquidas suspendidas en el aire que tienen un diámetro aerodinámico igual o menor a 10 micrómetros (μm). Al igual que las PM2.5 (partículas menores a 2.5 micras), las PM10 son un tipo de contaminante atmosférico particulado. Aunque son un poco más grandes que las PM2.5, todavía son lo suficientemente pequeñas como para permanecer suspendidas en el aire durante largos períodos de tiempo y penetrar en el sistema respiratorio cuando se inhalan (Rodríguez, 2017).

Fuentes de emisión:

Las PM10 pueden tener diversas fuentes, tanto naturales como antropogénicas (causadas por actividades humanas), que incluyen:

- Polvo y partículas naturales: Las partículas PM10 pueden generarse a partir de fuentes naturales como el polvo del suelo, la erosión del suelo, las tormentas de arena y las cenizas volcánicas.
- Combustión de combustibles fósiles: La quema de carbón, petróleo y gas natural en centrales eléctricas, vehículos de motor y calentadores también puede liberar PM10.
- Emisiones industriales: Las actividades industriales, como la fundición, la fabricación y la construcción, pueden ser fuentes significativas de PM10.
- Emisiones de agricultura: La agricultura, a través de prácticas como la quema de rastrojos y la aplicación de fertilizantes, puede contribuir a la liberación de partículas PM10.
- Actividades de construcción: Las operaciones de construcción, como la demolición y el movimiento de tierras, pueden generar PM10 en el aire.

Impactos en la salud:

La exposición a concentraciones elevadas de PM10 puede tener efectos adversos en la salud humana, que incluyen:

- Irritación respiratoria: Las PM10 pueden irritar las vías respiratorias, causando tos, congestión nasal, dificultad para respirar y agravamiento de enfermedades respiratorias preexistentes, como el asma y la bronquitis (Ibarlucía, 2019).
- Problemas cardiovasculares: La exposición a PM10 se ha relacionado con un mayor riesgo de problemas cardiovasculares, como enfermedades del corazón y accidentes cerebrovasculares.
- Efectos en la salud infantil: Los niños pueden ser particularmente vulnerables a los efectos de PM10 debido a sus sistemas respiratorios aún en desarrollo.

Regulación y monitoreo:

Al igual que con las PM2.5, muchos países tienen regulaciones y estándares de calidad del aire para limitar las concentraciones de PM10 en el ambiente. El monitoreo constante de la calidad del aire se realiza para evaluar los niveles de PM10 y tomar medidas adecuadas cuando se superan los límites permitidos.

El control de las emisiones de PM10 y la adopción de prácticas más limpias en industrias, construcción y agricultura son esenciales para reducir los efectos negativos en la salud humana y el medio ambiente asociados con estas partículas. También es importante tomar medidas para reducir la exposición personal, como evitar actividades al aire libre durante episodios de alta concentración de PM10 y utilizar máscaras respiratorias en situaciones de riesgo.

2.2.6. Dióxido de azufre (SO₂)

El dióxido de azufre (SO₂) es un gas incoloro con un olor característico a huevo podrido. Se forma a partir de la combustión de compuestos que contienen azufre, como el carbón y el petróleo, y es uno de los contaminantes atmosféricos más comunes. A continuación, se presentan algunos aspectos clave relacionados con el dióxido de azufre (SO₂) (Comunidad de Madrid,, 2021):

Fuentes de emisión:

Combustión de combustibles fósiles: La quema de carbón, petróleo y gas natural en centrales eléctricas, plantas industriales y vehículos de motor es una de las principales fuentes de emisión de SO₂.

- **Actividades industriales:** Las industrias como la refinación de petróleo, la fabricación de papel y la producción de metales pueden liberar SO₂ como subproducto de sus procesos.
- **Erupciones volcánicas:** Las erupciones volcánicas pueden liberar grandes cantidades de SO₂ en la atmósfera.

Impactos en la salud y el medio ambiente:

- **Efectos en la salud humana:** La exposición al SO₂ puede causar una serie de problemas de salud, incluyendo irritación de las vías respiratorias, tos, dificultad para respirar y exacerbación de enfermedades respiratorias preexistentes, como el asma y la bronquitis.
- **Formación de lluvia ácida:** El SO₂, cuando se combina con otros contaminantes atmosféricos, como el óxido de nitrógeno (NO_x), puede contribuir a la formación de lluvia ácida. Esto puede dañar los suelos, los cuerpos de agua y los ecosistemas, afectando la biodiversidad y la salud de los ecosistemas acuáticos y terrestres.
- **Daño a materiales y edificios:** El SO₂ también puede corroer metales y dañar edificios y monumentos históricos.

Control y mitigación:

La reducción de las emisiones de dióxido de azufre es fundamental para prevenir sus impactos negativos, las estrategias para controlar las emisiones de SO₂ incluyen:

- **Tecnologías de control de la contaminación:** El uso de dispositivos de control de la contaminación, como precipitadores electrostáticos y scrubbers (depuradores), en plantas de energía e instalaciones industriales puede capturar el SO₂ antes de que sea liberado a la atmósfera (Chiluisa, 2023).
- **Desulfuración de combustibles:** La reducción del contenido de azufre en los combustibles utilizados en centrales eléctricas y vehículos es una medida efectiva para reducir las emisiones de SO₂.
- **Regulación gubernamental:** Muchos países tienen regulaciones y estándares de calidad del aire que limitan las emisiones de SO₂. Estos estándares establecen límites máximos permitidos para las emisiones de SO₂ y requieren que las empresas adopten tecnologías de control para cumplir con estos límites.

La reducción de las emisiones de SO₂ es esencial para mejorar la calidad del aire y prevenir problemas de salud y ambientales asociados con este contaminante atmosférico.

2.2.7. Dirección del Viento (DIR)

La dirección del viento (DIR) es un término que se utiliza en meteorología y navegación para describir la dirección desde la cual sopla el viento. Se expresa en grados, generalmente medidos en sentido horario a partir del norte geográfico, de manera que el norte es 0 grados, el este es 90 grados, el sur es 180 grados y el oeste es 270 grados. La dirección del viento se utiliza para indicar de dónde proviene el viento en relación con un punto de referencia y es una parte importante de la información meteorológica utilizada en pronósticos y para la navegación en mar, aire y tierra (Chicón, 2023).

2.2.8. Datos Humedad relativa (HUM)

La humedad relativa (HUM) es una medida que describe la cantidad de humedad presente en el aire en relación con la máxima cantidad de humedad que el aire podría contener a una temperatura y presión dadas. La humedad relativa se expresa generalmente en porcentaje (%). Aquí hay algunos ejemplos de datos de humedad relativa en diferentes contextos (Ibarlucía, 2019):

- **Clima y pronóstico del tiempo:** Los informes meteorológicos suelen incluir la humedad relativa como parte de la información sobre el tiempo. Por ejemplo, un informe meteorológico podría indicar que la humedad relativa es del 60%, lo que significa que el aire contiene el 60% de la cantidad máxima de humedad que podría contener a esa temperatura y presión.
- **Control de clima interior:** En entornos interiores, como hogares y oficinas, se mide la humedad relativa para garantizar un nivel de confort adecuado. Por lo general, se busca mantener la humedad relativa en un rango cómodo, que generalmente oscila entre el 30% y el 60%, para evitar problemas como el aire seco o la condensación excesiva.
- **Agricultura:** La humedad relativa es importante en la agricultura, ya que puede afectar el crecimiento de cultivos y la salud de las plantas. Un nivel adecuado de humedad relativa es esencial para la fotosíntesis y el desarrollo de las plantas.
- **Navegación y aviación:** En la navegación marítima y la aviación, conocer la humedad relativa es esencial para calcular la densidad del aire, lo que a su vez afecta la eficiencia de los motores y la altitud a la que una aeronave puede volar.
- **Ciencias ambientales:** En estudios de calidad del aire y ciencias ambientales, la humedad relativa se mide para comprender cómo influye en la formación de contaminantes atmosféricos y la dispersión de partículas.
- **Procesos industriales:** En ciertas industrias, como la fabricación de productos farmacéuticos o electrónicos, el control preciso de la humedad relativa es esencial para garantizar la calidad de los productos.

La humedad relativa es una medida importante para comprender cómo interactúa la humedad en el aire con diversas actividades y procesos en la vida cotidiana y en campos especializados. Cambios en la humedad relativa pueden influir en la salud, el confort y la seguridad, así como en la eficiencia de las operaciones y procesos industriales.

2.2.9. Datos radiación ultravioleta (IUV)

La radiación ultravioleta (UV) es una forma de radiación electromagnética emitida por el sol y, en menor medida, por otras fuentes naturales y artificiales. La exposición excesiva a la radiación ultravioleta puede tener efectos dañinos en la piel, los ojos y la salud en general. Para medir la intensidad de la radiación ultravioleta y evaluar su impacto en la salud humana, se utiliza el Índice

Ultravioleta (IUV) o Índice UV. Este índice proporciona información sobre la fuerza de la radiación UV en un lugar y momento específicos.

- Informes meteorológicos: En muchos informes meteorológicos o aplicaciones de pronóstico del tiempo, se incluye el Índice UV como parte de la información relacionada con el clima. Por ejemplo, un informe podría indicar que el Índice UV es 8, lo que significa que la radiación UV es de intensidad "muy alta".
- Pantallas y aplicaciones móviles: Algunas pantallas digitales en áreas públicas o aplicaciones móviles brindan información en tiempo real sobre el Índice UV en una ubicación específica. Esto puede ayudar a las personas a tomar decisiones sobre la exposición al sol.
- Consejos de protección solar: Junto con la información del Índice UV, es común proporcionar consejos sobre cómo protegerse del sol en función de la intensidad de la radiación UV. Por ejemplo, en días con un Índice UV alto, se aconseja usar protector solar, gafas de sol y ropa protectora.
- Monitoreo de la exposición: Algunas personas, especialmente aquellas con mayor riesgo de daño por radiación UV, como personas con piel sensible o antecedentes de cáncer de piel, pueden usar el Índice UV para monitorear su exposición y tomar precauciones adicionales cuando el índice es alto.

El Índice UV generalmente se escala en números enteros, donde valores más bajos indican una radiación UV más débil y valores más altos indican una radiación UV más intensa. Los rangos típicos del Índice UV son (Chicón, 2023):

- Bajo: 0-2
- Moderado: 3-5
- Alto: 6-7
- Muy Alto: 8-10
- Extremadamente Alto: 11 o más

Es importante prestar atención al Índice UV y tomar medidas adecuadas para protegerse cuando la radiación UV es alta, ya que la exposición excesiva al sol puede aumentar el riesgo de daño en la piel, como quemaduras solares y cáncer de piel, así como daños en los ojos.

2.2.10. Datos precipitación (LLU)

La precipitación (generalmente abreviada como "LLU" o "Precip.") se refiere a cualquier forma de agua que cae de la atmósfera hacia la superficie terrestre. Esta forma de agua puede incluir lluvia, nieve, granizo, aguanieve y llovizna, entre otras. La precipitación es un componente clave de la hidrología y del ciclo del agua en la Tierra, y su cantidad y tipo pueden variar según la ubicación geográfica y las condiciones meteorológicas (Malagón, 2011).

- **Informes meteorológicos:** Los informes meteorológicos y las aplicaciones de pronóstico del tiempo a menudo proporcionan información sobre la precipitación actual, así como pronósticos de precipitación futura. Los datos de precipitación se presentan en unidades de medida como milímetros (mm) o pulgadas (in) (Narváez, 2021).
- **Estaciones meteorológicas:** Las estaciones meteorológicas terrestres y las estaciones de observación meteorológica miden la cantidad de precipitación que cae en un área específica. Estas estaciones pueden utilizar instrumentos como pluviómetros o cuentagotas para recopilar datos precisos sobre la precipitación.
- **Radars meteorológicos:** Los radares meteorológicos se utilizan para monitorear la precipitación en tiempo real y en un área más amplia. Los datos de radar pueden mostrar la intensidad y la ubicación de la precipitación en un mapa, lo que ayuda a los meteorólogos a rastrear tormentas y sistemas de mal tiempo (Organización Mundial de la Salud, 2019).
- **Mapas de precipitación:** Los mapas de precipitación son representaciones visuales de la cantidad de precipitación que ha caído en una región específica durante un período de tiempo determinado. Estos mapas pueden ser generados por agencias meteorológicas y organizaciones de seguimiento climático.
- **Registros climáticos:** Los registros climáticos a largo plazo incluyen datos de precipitación que se han recopilado durante décadas o incluso siglos en algunas ubicaciones. Estos registros ayudan a los científicos a comprender los patrones climáticos y las tendencias a lo largo del tiempo.

Los datos de precipitación son fundamentales para muchas aplicaciones, como la gestión del agua, la agricultura, la predicción de inundaciones, la climatología y la planificación urbana. También son cruciales para evaluar y abordar desafíos relacionados con el cambio climático, ya que las

variaciones en los patrones de precipitación pueden tener un impacto significativo en los ecosistemas y en la disponibilidad de agua dulce.

2.2.11. Presión barométrica (PRE)

La presión barométrica, a menudo abreviada como "PRE" o simplemente "presión", se refiere a la presión atmosférica en un lugar y momento específicos. Es la fuerza ejercida por la columna de aire que se encuentra sobre una unidad de superficie y se mide en unidades como hectopascales (hPa) o milibares (mb). La presión barométrica es una medida importante en meteorología y se utiliza en la predicción del tiempo y para comprender las condiciones atmosféricas (Aránguez, 2019).

Aquí hay ejemplos de cómo se presentan los datos de presión barométrica:

- **Informes meteorológicos:** Los informes meteorológicos suelen incluir la presión barométrica como una de las medidas clave. Se puede expresar en hectopascales (hPa) o milibares (mb) y se muestra junto con otros datos meteorológicos como temperatura, humedad y velocidad del viento.
- **Gráficos meteorológicos:** Los mapas meteorológicos utilizan isotermas (líneas de igual temperatura) e isobaras (líneas de igual presión barométrica) para representar las condiciones atmosféricas a nivel regional o global. Estos gráficos muestran patrones de alta y baja presión, lo que ayuda a los meteorólogos a prever el tiempo.
- **Instrumentos de medición:** Las estaciones meteorológicas y los instrumentos de medición en estaciones de observación meteorológica miden la presión barométrica. Los barómetros son dispositivos diseñados específicamente para medir la presión atmosférica y se utilizan en aplicaciones meteorológicas y de navegación.
- **Pronósticos meteorológicos:** Los pronósticos meteorológicos utilizan datos de presión barométrica para predecir el comportamiento futuro de las condiciones atmosféricas. Los cambios en la presión barométrica pueden indicar la llegada de sistemas meteorológicos, como frentes fríos o sistemas de baja presión.

La presión barométrica varía según la ubicación geográfica y las condiciones meteorológicas. Las áreas de alta presión suelen estar asociadas con condiciones más estables y cielos despejados,

mientras que las áreas de baja presión a menudo indican la posibilidad de condiciones climáticas inestables, como lluvia o tormentas (Ibarlucía, 2019).

El seguimiento de la presión barométrica es importante para comprender y predecir las condiciones meteorológicas y para diversas aplicaciones, como la navegación, la aviación y la planificación de actividades al aire libre.

2.2.12. Temperatura media (TMP)

La temperatura media se refiere al valor promedio de la temperatura en un lugar geográfico durante un período de tiempo específico. Esta medida es fundamental para entender las condiciones climáticas en una región determinada y se utiliza en una variedad de contextos, desde el pronóstico del tiempo hasta la investigación climática y la toma de decisiones en agricultura, energía y diseño de edificios, entre otros (Narváez, 2021).

Aquí hay algunos aspectos clave relacionados con la temperatura media:

- **Cálculo de la temperatura media:** Para calcular la temperatura media, se suman todas las temperaturas registradas en un período determinado (por ejemplo, durante un mes) y se dividen por el número de observaciones. Esto proporciona un valor promedio que refleja las condiciones térmicas durante ese período.
- **Unidades de medida:** La temperatura se mide comúnmente en grados Celsius ($^{\circ}\text{C}$) o grados Fahrenheit ($^{\circ}\text{F}$), dependiendo de la región. La temperatura media se expresará en la misma unidad que las observaciones individuales.
- **Aplicaciones climáticas:** La temperatura media es esencial para la clasificación de los climas, ya que se utiliza para determinar si una región es tropical, templada, fría, entre otras. También se utiliza en la identificación de estaciones del año, como el promedio de temperaturas en verano, primavera, otoño e invierno (Narváez, 2021).
- **Cambios climáticos:** Los científicos y climatólogos utilizan datos de temperatura media a lo largo del tiempo para analizar patrones climáticos y detectar tendencias de cambio climático, como el aumento de las temperaturas promedio globales.
- **Planificación agrícola:** La temperatura media es un factor crítico para la agricultura, ya que influye en el crecimiento de cultivos y la selección de variedades adecuadas. Los

agricultores utilizan los registros de temperatura media para planificar las épocas de siembra y cosecha.

- Diseño de edificios y energía: En la arquitectura y la ingeniería, la temperatura media se tiene en cuenta al diseñar edificios y sistemas de calefacción y refrigeración para garantizar el confort térmico y la eficiencia energética.

2.2.13. Velocidad del viento (VEL)

La velocidad del viento (VEL) se refiere a la rapidez o la velocidad a la que el viento se mueve en un lugar y momento específicos. Se mide en unidades como metros por segundo (m/s), kilómetros por hora (km/h), millas por hora (mph) o nudos (una unidad de velocidad náutica). La velocidad del viento es una medida fundamental en meteorología y se utiliza en una variedad de aplicaciones, desde la navegación y la aviación hasta la generación de energía eólica y la predicción del tiempo (Paredes, 2020).

- Métodos de medición: La velocidad del viento se mide utilizando instrumentos llamados anemómetros. Estos dispositivos pueden variar desde simples copas giratorias hasta sensores electrónicos avanzados. Los anemómetros registran la velocidad del viento en tiempo real y envían datos a estaciones meteorológicas o sistemas de control.
- Factores influenciadores: La velocidad del viento puede variar significativamente según la ubicación geográfica y las condiciones climáticas. Los factores que influyen en la velocidad del viento incluyen la altitud, la topografía, la proximidad a cuerpos de agua, la presión atmosférica y la existencia de sistemas meteorológicos, como frentes y tormentas.

Nomenclatura de velocidades del viento

- Calma: Velocidad del viento menor a 1 m/s o 2 mph.
- Brisa ligera: Velocidad del viento entre 1 m/s (2 mph) y 5 m/s (11 mph).
- Brisa moderada: Velocidad del viento entre 5 m/s (11 mph) y 8 m/s (18 mph).
- Brisa fresca: Velocidad del viento entre 8 m/s (18 mph) y 12 m/s (27 mph).
- Viento fuerte: Velocidad del viento entre 12 m/s (27 mph) y 20 m/s (45 mph).
- Viento muy fuerte: Velocidad del viento entre 20 m/s (45 mph) y 25 m/s (56 mph).
- Vendaval: Velocidad del viento mayor a 25 m/s (56 mph).

Aplicaciones prácticas: La velocidad del viento es crucial en varias industrias y aplicaciones, como la navegación marítima y aérea, la generación de energía eólica, la planificación urbana y la predicción del tiempo. También es importante para determinar factores como la sensación térmica y la velocidad del viento en eventos climáticos extremos como huracanes y tornados.

2.3 Sensores de medición

Los sensores de medición son dispositivos diseñados para detectar y cuantificar una amplia variedad de propiedades físicas, químicas o ambientales. Estos sensores convierten una magnitud específica en una señal eléctrica o digital que puede ser procesada y utilizada para monitorear, controlar o medir una condición o parámetro en tiempo real. Los sensores son fundamentales en muchas aplicaciones y campos, incluyendo la electrónica, la automatización industrial, la medicina, la ciencia ambiental, la meteorología y muchos otros (Placeres, 2020).

- Sensor de temperatura: Mide la temperatura ambiente o de un objeto y puede utilizar tecnologías como termistores, termopares o sensores de resistencia de platino (RTD).
- Sensor de humedad: Mide la cantidad de humedad presente en el aire u otros medios, basándose en capacitancia, resistencia o técnicas ópticas.
- Sensor de presión: Mide la presión de fluidos o gases e incluye sensores piezorresistivos, de diafragma y de tubo Bourdon.
- Sensor de luz: Detecta la intensidad de la luz visible o infrarroja, empleando fotorresistores, fotodiodos o fototransistores.
- Sensor de sonido (micrófono): Convierte las ondas de sonido en señales eléctricas para medir el sonido o el ruido ambiental.
- Sensor de movimiento (acelerómetro): Mide la aceleración en una o más direcciones, utilizándose en dispositivos de navegación y detección de impactos.
- Sensor de proximidad: Detecta la proximidad de un objeto sin contacto físico, utilizando sensores ultrasónicos y de infrarrojos.
- Sensor de gas: Mide la concentración de un gas específico en el aire y se utiliza en aplicaciones de calidad del aire y detección de gases tóxicos.
- Sensor de pH: Mide la acidez o alcalinidad de una solución, aplicándose en química analítica y control de calidad.

- Sensor de flujo: Mide la velocidad o el caudal de un fluido, como líquidos o gases, siendo esencial en control de procesos industriales y monitoreo ambiental.
- Sensor de GPS: Utiliza señales satelitales para determinar la ubicación geográfica precisa, ampliamente utilizado en sistemas de navegación y rastreo.
- Sensor de imagen (Cámara): Captura imágenes visuales o térmicas para diversas aplicaciones, desde fotografía digital hasta imágenes médicas y vigilancia.

2.4 Machine Learning

El aprendizaje automático, o machine learning en inglés, es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas a partir de la experiencia y los datos, en lugar de ser programadas explícitamente para realizar una tarea. El aprendizaje automático se basa en la idea de que las computadoras pueden identificar patrones y tomar decisiones por sí mismas, a medida que se les proporciona más información y experiencia (Sanitaria, 2019).

Aquí hay algunos conceptos clave relacionados con el aprendizaje automático:

- Algoritmos de aprendizaje: Los algoritmos de aprendizaje son reglas y procedimientos matemáticos diseñados para aprender de los datos. Algunos ejemplos comunes de algoritmos de aprendizaje incluyen árboles de decisión, regresión lineal, redes neuronales, máquinas de vectores de soporte (SVM) y algoritmos de agrupamiento como k-means.
- Conjuntos de datos: Para entrenar y evaluar modelos de aprendizaje automático, se utilizan conjuntos de datos que contienen ejemplos etiquetados o datos históricos. Estos datos se dividen típicamente en conjuntos de entrenamiento (para el aprendizaje) y conjuntos de prueba (para la evaluación del rendimiento del modelo) (Malagón, 2011).
- Supervisado vs. no supervisado: En el aprendizaje supervisado, los modelos se entrenan utilizando ejemplos etiquetados, lo que significa que se conoce la respuesta correcta. En el aprendizaje no supervisado, los modelos encuentran patrones y estructuras en los datos sin etiquetar.
- Aprendizaje profundo (Deep Learning): Es una subdisciplina del aprendizaje automático que se centra en el uso de redes neuronales profundas para tareas de procesamiento de datos complejas, como reconocimiento de imágenes, procesamiento de lenguaje natural y análisis de datos secuenciales.

Aplicaciones: El aprendizaje automático se utiliza en una amplia variedad de aplicaciones, como reconocimiento facial, recomendaciones de productos en línea, diagnóstico médico, conducción autónoma, traducción de idiomas, detección de fraudes y muchas otras áreas.

Evaluación del modelo: Para evaluar la calidad de un modelo de aprendizaje automático, se utilizan métricas como la precisión, el error cuadrático medio (MSE), el coeficiente de determinación (R-squared), entre otros, dependiendo del tipo de tarea y del conjunto de datos.

Entrenamiento y ajuste de hiperparámetros: Los modelos de aprendizaje automático requieren un proceso de entrenamiento, donde los algoritmos ajustan sus parámetros para adaptarse a los datos de entrenamiento. Además, se pueden ajustar hiperparámetros, como tasas de aprendizaje y profundidad de los modelos, para mejorar el rendimiento.

2.4.1 Descripción e importancia de machine learning

El aprendizaje automático, o machine learning en inglés, es un enfoque de la inteligencia artificial (IA) que se centra en desarrollar algoritmos y modelos que permiten a las computadoras aprender y tomar decisiones basadas en datos y experiencia, en lugar de depender de una programación explícita. En lugar de ser programadas de manera rígida para realizar tareas específicas, las máquinas de aprendizaje automático pueden analizar datos, identificar patrones y tomar decisiones adaptativas (Sanitaria, 2019).

Los modelos de aprendizaje automático se entrenan utilizando conjuntos de datos que contienen ejemplos etiquetados o datos históricos. Estos modelos se ajustan a los datos durante el proceso de entrenamiento y luego pueden aplicarse a nuevos datos para hacer predicciones o tomar decisiones.

2.4.2 Describir el Pipeline de un proceso de machine learning

El pipeline de un proceso de machine learning se refiere a la secuencia de pasos organizados y conectados que se siguen para desarrollar y entrenar un modelo de aprendizaje automático. Cada paso en el pipeline cumple una función específica y contribuye al proceso general de construcción de un modelo que pueda realizar predicciones o tomar decisiones basadas en datos. A continuación, se describe un pipeline típico de un proceso de machine learning (Chicón, 2023):

- Recopilación de Datos.
- Preprocesamiento de Datos.
- División del Conjunto de Datos.

- Selección del Modelo.
- Entrenamiento del Modelo.
- Validación del Modelo.
- Ajuste de Hiperparámetros.
- Implementación del Modelo.
- Evaluación Continua:

2.4.3 Descripción de Deep learning

El aprendizaje profundo, conocido como deep learning en inglés, es una subdisciplina del aprendizaje automático (machine learning) que se centra en la construcción y entrenamiento de redes neuronales artificiales profundas. Las redes neuronales profundas son modelos computacionales inspirados en la estructura y el funcionamiento del cerebro humano, específicamente en las redes de neuronas biológicas interconectadas. El aprendizaje profundo ha ganado un gran interés y popularidad debido a su capacidad para resolver problemas complejos de procesamiento de datos, como reconocimiento de imágenes, procesamiento de lenguaje natural, juegos y más (Gallegos, 2018).

- Redes neuronales artificiales (RNA): En el aprendizaje profundo, se utilizan redes neuronales artificiales para modelar datos y aprender representaciones abstractas de estos datos. Estas redes están compuestas por múltiples capas de neuronas artificiales interconectadas.
- Capas ocultas: Una característica distintiva del aprendizaje profundo es la presencia de múltiples capas ocultas en una red neuronal, lo que le da profundidad al modelo. Cuantas más capas tiene una red, más "profunda" es. Cada capa oculta realiza operaciones matemáticas para aprender características cada vez más abstractas de los datos.
- Conexiones ponderadas: Las conexiones entre las neuronas de una red neuronal tienen pesos asociados. Estos pesos se ajustan durante el entrenamiento de la red para que la red pueda aprender y adaptarse a los datos de entrada .

Entrenamiento y Aprendizaje:

El proceso de entrenamiento en el aprendizaje profundo implica los siguientes pasos:

- Inicialización de pesos: Los pesos de la red se inicializan aleatoriamente.

- Pase hacia adelante (Forward Pass): Los datos de entrada se propagan a través de la red desde la capa de entrada hasta la capa de salida. Durante este proceso, se realizan cálculos en cada capa para generar predicciones.
- Cálculo del error: Se compara la salida de la red con las respuestas reales y se calcula un error, el objetivo es minimizar este error.
- Retropropagación (Backpropagation): Se retropropagan los errores desde la capa de salida hacia las capas ocultas. Los gradientes se utilizan para ajustar los pesos de las conexiones en la dirección que reducirá el error.
- Actualización de pesos: Los pesos de la red se actualizan utilizando algoritmos de optimización, como el descenso de gradiente estocástico (SGD), para minimizar el error.
- Iteración: Estos pasos se repiten iterativamente en múltiples épocas hasta que el modelo converja y alcance un rendimiento satisfactorio en los datos de entrenamiento.

Aplicaciones del aprendizaje profundo:

- El aprendizaje profundo ha demostrado ser eficaz en una amplia variedad de aplicaciones, incluyendo:
- Reconocimiento de imágenes: Identificación y clasificación de objetos en imágenes y videos (Chicón, 2023).
- Procesamiento de lenguaje natural (NLP): Traducción de idiomas, generación de texto, análisis de sentimientos y chatbots.
- Reconocimiento de voz: Transcripción de voz a texto y sistemas de reconocimiento de comandos de voz.
- Juegos: Aprendizaje profundo aplicado a juegos como ajedrez, Go y videojuegos.
- Salud: Diagnóstico médico, análisis de imágenes médicas y predicción de enfermedades.
- Automatización industrial: Control de procesos y detección de fallos en la producción industrial.
- Conducción autónoma: Navegación y toma de decisiones en vehículos autónomos.

2.5 Estudios previos

Título: "Predicción de la Calidad del Aire en Madrid mediante Modelos de Machine Learning"

Este estudio se centró en el desarrollo de modelos de aprendizaje automático para predecir la calidad del aire en la ciudad de Madrid. Se utilizaron datos históricos de calidad del aire, información meteorológica y datos de tráfico para entrenar y evaluar los modelos. Los resultados proporcionaron información valiosa para la gestión de la calidad del aire en la capital española (Aránguez, 2019).

Título: "Modelo de Predicción de Contaminación Atmosférica para Santiago de Chile"

En este estudio, se desarrolló un modelo de predicción de la contaminación atmosférica específicamente para la ciudad de Santiago. Se utilizaron datos de estaciones de monitoreo de calidad del aire y datos meteorológicos para predecir los niveles de contaminantes atmosféricos. El objetivo era mejorar la capacidad de pronóstico y la toma de decisiones en relación con la contaminación del aire en la capital chilena (Gallegos, 2018).

Título: "Predicción de la Calidad del Aire en Ciudad de México utilizando Redes Neuronales Artificiales"

En esta investigación, se emplearon redes neuronales artificiales para predecir la calidad del aire en la Ciudad de México. Se utilizaron datos históricos de calidad del aire, condiciones meteorológicas y factores geográficos para desarrollar el modelo. Los resultados contribuyeron a la gestión de la calidad del aire en una de las ciudades más pobladas del mundo (Ibarlucía, 2019).

Título: "Modelo de Predicción de Contaminantes Atmosféricos en Bogotá: Un Enfoque Multivariado"

En este estudio, se desarrolló un modelo de predicción de contaminantes atmosféricos en Bogotá que tuvo en cuenta múltiples variables, como datos de calidad del aire, factores geográficos y condiciones meteorológicas. El objetivo era proporcionar pronósticos precisos de la calidad del aire en la capital colombiana (Comunidad de Madrid,, 2021).

Título: "Predicción de la Contaminación del Aire en el Área Metropolitana de Buenos Aires utilizando Machine Learning"

Este estudio se enfocó en el área metropolitana de Buenos Aires y utilizó técnicas de aprendizaje automático para predecir la contaminación del aire. Los datos de calidad del aire y las condiciones meteorológicas se utilizaron para desarrollar el modelo de predicción (Aránguez, 2019).

Dentro del presente trabajo de investigación se halla el valor agregado la experimentación y comparación con más variedad de algoritmos de aprendizaje automático, esto permite hallar diversas soluciones a un mismo problema que se acople de mejor manera al problema que se trata de abordar.

Además de brindar ciertas opciones en cuanto a la selección de un modelo de caja negra en el caso de modelos de Deep learning o de caja blanca en el caso de algoritmos de machine learning para mejorar el entendimiento y comprensión de los resultados obtenidos.

3. Metodología

3.1. Tipo de estudio

En el tipo de estudio, se especificará el tipo de estudio que se llevará a cabo, dado el alcance de la investigación sobre contaminación, se propone un enfoque mixto que combina elementos exploratorios, descriptivos y correlacionales. La elección de este enfoque se justifica debido a la complejidad de la contaminación y su variabilidad en el tiempo y el espacio. El enfoque exploratorio permitirá identificar nuevas tendencias, el enfoque descriptivo proporcionará una imagen detallada de la situación actual y el enfoque correlacional ayudará a comprender las relaciones entre las variables.

3.2. Descripción de datos y fuente de información

En este apartado, se detallará cómo se recopilarán los datos necesarios para la investigación sobre la contaminación. Los datos se obtendrán principalmente de fuentes confiables, como la Agencia Metropolitana de Control del Aire de Quito (AMCA), estaciones de monitoreo ambiental y bases de datos públicas. Los datos serán de diversas naturalezas, incluyendo mediciones de contaminantes atmosféricos, datos meteorológicos, geográficos y demográficos. Se prestará especial atención a la calidad, consistencia y temporalidad de los datos.

Para llevar a cabo esta investigación sobre la contaminación del aire en la ciudad de Quito, se requerirá una fuente de datos confiable y variada. La calidad de los datos y su disponibilidad son fundamentales para obtener resultados precisos y representativos. La metodología de recopilación de datos se divide en varias etapas (Gallegos, 2018):

Fuentes de datos confiables:

- **Agencia Metropolitana de Control del Aire de Quito (AMCA):** Esta será una de las fuentes principales de datos. La AMCA monitorea y registra datos de calidad del aire en Quito utilizando estaciones de monitoreo ambiental en diferentes ubicaciones de la ciudad. Estos datos incluyen mediciones de contaminantes atmosféricos, como partículas PM2.5 y PM10, óxidos de nitrógeno (NOx), dióxido de azufre (SO2), monóxido de carbono (CO), entre otros.
- **Estaciones de Monitoreo Ambiental:** Además de la AMCA, otras estaciones de monitoreo ambiental, tanto gubernamentales como privadas, pueden proporcionar datos relevantes. Estas estaciones pueden estar ubicadas en áreas específicas de interés, como zonas industriales o urbanas, y proporcionar mediciones detalladas.
- **Bases de datos públicas:** Se explorarán bases de datos públicas que contengan información relevante sobre la contaminación del aire en Quito. Esto podría incluir datos meteorológicos históricos, datos geográficos (como ubicaciones de fuentes de contaminación) y datos demográficos que ayuden a contextualizar la investigación.

Naturaleza de los datos:

- **Contaminantes atmosféricos:** Los datos incluirán mediciones de contaminantes atmosféricos, como concentraciones de partículas en suspensión (PM2.5 y PM10), gases contaminantes (NOx, SO2, CO, O3), y otros indicadores clave de la calidad del aire.
- **Datos meteorológicos:** Se incluirán datos meteorológicos relevantes, como temperatura, humedad relativa, velocidad y dirección del viento, que pueden influir en los patrones de contaminación del aire.
- **Datos geográficos y demográficos:** Los datos geográficos, como la ubicación de las fuentes de contaminación, la topografía de la ciudad y la densidad de población, proporcionarán información valiosa sobre las características locales que pueden afectar la contaminación del aire en Quito.

Calidad, consistencia y temporalidad de los datos:

- **Calidad de los datos:** Se realizará un control de calidad riguroso para garantizar la fiabilidad de los datos. Esto incluirá la detección y corrección de valores atípicos, la verificación de la precisión de las mediciones y la garantía de la coherencia en los registros.
- **Consistencia:** Se verificará la consistencia de los datos a lo largo del tiempo y entre diferentes fuentes. La falta de coherencia podría ser indicativa de problemas en las mediciones o en la recopilación de datos.
- **Temporalidad:** Se considerará la temporalidad de los datos, asegurándose de que abarquen un período suficientemente largo para analizar tendencias a lo largo de los años y las estaciones.

3.3. Descripción de modelos de machine learning supervisado

En este apartado se explicarán los modelos de machine learning supervisado que se utilizarán para abordar el problema de la contaminación. Se seleccionarán algoritmos apropiados para la tarea, como regresión lineal, regresión logística, máquinas de soporte vectorial y redes neuronales, debido a su capacidad para manejar datos complejos y relaciones no lineales. Se justificará cada elección en función de su adecuación para el problema y se describirá cómo se aplicarán.

Regresión Lineal: La regresión lineal será uno de los modelos utilizados debido a su simplicidad y capacidad para establecer relaciones lineales entre las variables predictoras y la variable objetivo. Se aplicará cuando se sospeche que existen relaciones lineales entre ciertas variables (por ejemplo, la influencia de la temperatura en la concentración de contaminantes).

Máquinas de Soporte Vectorial (SVM): Las SVM son útiles cuando se busca clasificar y predecir datos no lineales. Se utilizarán para abordar posibles relaciones no lineales en los datos, permitiendo la identificación de patrones complejos en la contaminación del aire en Quito.

Vecinos más cercanos (KNN): El algoritmo K-Nearest Neighbors es un método de aprendizaje supervisado utilizado tanto para clasificación como para regresión. Su enfoque fundamental es simple: predice el valor de una nueva instancia basándose en la mayoría de las clases (en clasificación) o el promedio de los valores (en regresión) de sus K vecinos más cercanos en el espacio de características, este último es el que se usa en el que se aborda en el presente trabajo ya que la variable a predecir es el número de partículas en el aire de Quito.

Árbol de predicción: Un "árbol de predicción" generalmente se refiere a un tipo específico de modelo de aprendizaje supervisado llamado "árbol de decisión". Un árbol de decisión es una estructura de árbol que se utiliza para tomar decisiones basadas en condiciones sobre las características de los datos. Puede ser utilizado tanto para problemas de clasificación como de regresión.

Redes Neuronales Artificiales (ANN): Las redes neuronales serán una parte esencial de la metodología, especialmente si se busca capturar relaciones no lineales complejas y patrones en los datos. Se implementarán redes neuronales profundas para aprovechar al máximo su capacidad de aprendizaje profundo y adaptación a la complejidad del problema de la contaminación.

Justificación de la selección:

Regresión Lineal: Se justifica su uso en el contexto de la contaminación del aire debido a la posibilidad de que algunas relaciones sean lineales o aproximadamente lineales. Esto permitirá una interpretación más clara de la influencia de ciertas variables predictoras.

Máquinas de Soporte Vectorial (SVM): Las SVM son ideales para detectar relaciones no lineales y separar datos en clases o niveles de contaminación. Su capacidad para manejar datos complejos se adapta bien al problema de la contaminación.

Vecinos más cercanos (KNN): Los algoritmos de KNN sirven en el caso de hallar puntos o predecir observaciones en base a características similares de las observaciones más parecidas, lo que favorece la predicción en casos donde las observaciones tienen características similares a otras observaciones.

En el caso de los árboles de predicción, se realiza un proceso más discriminante al momento de realizar las predicciones, lo que permite segmentar resultados de una mejor manera para predecir la contaminación en el aire de Quito.

Redes Neuronales Artificiales (ANN): Se utilizarán redes neuronales debido a su capacidad para modelar relaciones no lineales altamente complejas y su capacidad para aprender automáticamente de los datos. Esto es esencial cuando se trata de comprender las interacciones entre múltiples variables que afectan la contaminación del aire.

Aplicación de Modelos:

Cada uno de los modelos seleccionados se aplicará en diferentes etapas del proyecto:

La regresión lineal se utilizará inicialmente para establecer relaciones básicas y comprender la importancia relativa de las variables predictoras.

Las SVM se aplicarán para explorar relaciones no lineales y clasificar niveles de contaminación del aire.

El modelo de KNN se utilizará para predecir valores que permitan conocer la calidad del aire en base a información de las observaciones más cercanas.

En cuanto al modelo de árbol de decisión, se procura tener un método de predicción, más jerárquico tomando en cuenta información más relevante o cementaciones más relevantes en las predicciones.

Las redes neuronales se utilizarán para modelar la complejidad subyacente en los datos y para realizar predicciones precisas y detalladas sobre la contaminación del aire en Quito.

3.4. Descripción de redes neuronales

Si se planea utilizar redes neuronales en la investigación, se proporcionará una descripción detallada de estos modelos. Se explicará cómo las redes neuronales artificiales funcionan como una analogía a las redes neuronales biológicas, y se describirá su arquitectura básica, incluyendo capas ocultas, funciones de activación y técnicas de entrenamiento, como el descenso del gradiente. Se mostrará cómo se adaptarán específicamente a la predicción de la contaminación del aire.

3.5. Métricas de evaluación del modelo

Se detallarán las métricas de evaluación del modelo que se utilizarán para medir el rendimiento de los algoritmos de machine learning y las redes neuronales. Se explicará cómo cada métrica, como el Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2), evalúa el rendimiento del modelo. También se proporcionará una guía sobre cómo interpretar los resultados y cómo estas métricas ayudarán a seleccionar el modelo más adecuado para la predicción de la contaminación del aire en Quito.

Las fórmulas utilizadas para la evaluación de los modelos se detallan a continuación.

MSE (Error Cuadrático Medio): El MSE mide el promedio de los errores al cuadrado entre las predicciones del modelo y los valores reales. Es particularmente sensible a errores grandes debido al término cuadrático.

$$MSE = \frac{\sum(\hat{y} - y)^2}{n}$$

RMSE (Raíz del Error Cuadrático Medio): El RMSE es la raíz cuadrada del MSE y proporciona una métrica en la misma escala que la variable objetivo original. Es interpretado fácilmente y menos sensible a valores atípicos que el MSE.

$$RMSE = \sqrt{\frac{\sum(\hat{y} - y)^2}{n}}$$

MAE (Error Absoluto Medio): El MAE mide el promedio de las diferencias absolutas entre las predicciones y los valores reales. Es menos sensible a valores atípicos que el MSE y RMSE.

$$MAE = \frac{\sum|\hat{y} - y|}{n}$$

Coefficiente de Determinación: El R² evalúa la proporción de la variabilidad de la variable dependiente explicada por el modelo. Un valor cercano a 1 indica un buen ajuste, mientras que cerca de 0 indica que el modelo no explica bien la variabilidad.

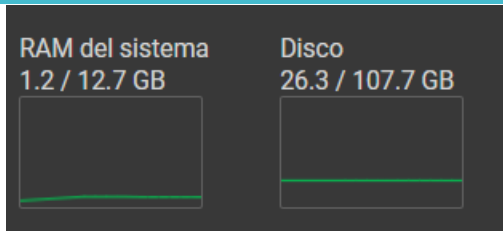
$$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$$

4. Análisis de resultados

Proceso de ETL de datos

Para el procesamiento y estimación de los modelos de predicción de la calidad del aire de la ciudad de Quito se utilizó la herramienta de Google Cloud dada la gran cantidad de datos, la máquina prestada por Google cuenta con 12.7 GB de memoria RAM y 107.7 GB de disco de almacenamiento que resulta suficiente para el presente trabajo (figura 1).

Figura 1. Características de la máquina en Google Cloud



Dado que los datos disponibles no se encuentran aún listos para poder realizar un análisis ya que se encuentran ubicados en diferentes repositorios de datos dentro de la URL <http://datosambiente.quito.gob.ec/> como se observa a continuación (figura 2).

Figura 2. Repositorio y distribución de datos disponibles

DATOS HISTORICOS REMMAQ DESCARGA 2004 - 2023

- [DATOS MONOXIDO CARBONO \(CO\)](#)
- [DATOS DIOXIDO DE NITROGENO \(NO2\)](#)
- [DATOS OZONO \(O3\)](#)
- [DATOS PARTICULAS MENORES A 2.5 MICROMETROS \(PM2.5\)](#)
- [DATOS PARTICULAS MENORES A 10 MICROMETROSS \(PM10\)](#)
- [DATOS DIOXIDO DE AZUFRE \(SO2\)](#)
- [DATOS DIRECCION DEL VIENTO \(DIR\)](#)
- [DATOS HUMEDAD RELATIVA \(HUM\)](#)
- [DATOS RADIACION ULTRAVIOLETA \(IUV\)](#)
- [DATOS PRECIPITACION \(LLU\)](#)
- [DATOS PRESION BAROMETRICA \(PRE\)](#)
- [DATOS RADIACION SOLAR \(RS\)](#)
- [DATOS TEMPERATURA MEDIA \(TMP\)](#)
- [DATOS VELOCIDAD DEL VIENTO \(VEL\)](#)

Para el proceso de importación, extracción y unión de datos es necesario primero descargar los archivos directamente de la página web para facilitar la ejecución del código por parte de cualquier usuario como se presenta en la figura 3.

Figura 3. Descarga de datos de contaminación

```
# Descarga de datos
!wget -q http://datosambiente.quito.gob.ec/datos/CO.rar
!wget -q http://datosambiente.quito.gob.ec/datos/NO2.rar
!wget -q http://datosambiente.quito.gob.ec/datos/O3.rar
!wget -q http://datosambiente.quito.gob.ec/datos/PM2.5.rar
!wget -q http://datosambiente.quito.gob.ec/datos/PM10.rar
!wget -q http://datosambiente.quito.gob.ec/datos/SO2.rar
!wget -q http://datosambiente.quito.gob.ec/datos/DIR.rar
!wget -q http://datosambiente.quito.gob.ec/datos/HUM.rar
!wget -q http://datosambiente.quito.gob.ec/datos/IUV.rar
!wget -q http://datosambiente.quito.gob.ec/datos/LLU.rar
!wget -q http://datosambiente.quito.gob.ec/datos/PRE.rar
!wget -q http://datosambiente.quito.gob.ec/datos/RS.rar
!wget -q http://datosambiente.quito.gob.ec/datos/TMP.rar
!wget -q http://datosambiente.quito.gob.ec/datos/VEL.rar
```

Una vez realizada la descarga de los datos dentro de la maquina prestada por Google, se crea un repositorio donde se almacenan los ficheros que se van a utilizar para el análisis.

Cabe recalcar que los datos que se descargan de cada uno de los links del URL presentado previamente es un archivo individual que contiene la información de una única variable por lo que existe un fichero por separado para las partes por millón de partículas de 2.5 micras (PM2.5), el monóxido de carbono (CO), el dióxido de nitrógeno (NO2), ozono (O3) y dióxido de azufre (SO2) y las variables atmosféricas para el análisis (figura 4).

Figura 4. Descompresión de datos en la carpeta XLSX

```
!unrar x 'CO.rar' /content/XLSX
!unrar x 'NO2.rar' /content/XLSX
!unrar x 'O3.rar' /content/XLSX
!unrar x 'PM2.5.rar' /content/XLSX
!unrar x 'PM10.rar' /content/XLSX
!unrar x 'SO2.rar' /content/XLSX
!unrar x 'DIR.rar' /content/XLSX
!unrar x 'HUM.rar' /content/XLSX
!unrar x 'IUV.rar' /content/XLSX
!unrar x 'LLU.rar' /content/XLSX
!unrar x 'PRE.rar' /content/XLSX
!unrar x 'RS.rar' /content/XLSX
!unrar x 'TMP.rar' /content/XLSX
!unrar x 'VEL.rar' /content/XLSX
```

Una vez realizada la descompresión de datos, se importa cada uno de los ficheros individuales para formar un único fichero de datos consolidado resultado de la unión de los 5 datasets importados (figura 5).

Figura 5. Importación individual de datos

```
import pandas as pd

# Factores ambientales
df_PM2_5 = pd.read_excel('PM2.5.xlsx', header=0)
df_O3 = pd.read_excel('O3.xlsx', header=0)
df_SO2 = pd.read_excel('SO2.xlsx', header=0)
df_NO2 = pd.read_excel('NO2.xlsx', header=0)
df_CO = pd.read_excel('CO.xlsx', header=0)

# Factores atmosféricos
df_HUM = pd.read_excel('HUM.xlsx')
df_VEL = pd.read_excel('VEL.xlsx')
df_LLU = pd.read_excel('LLU.xlsx')
df_IUV = pd.read_excel('IUV.xls')
df_RS = pd.read_excel('RS.xlsx')
df_TMP = pd.read_excel('TMP.xlsx')
```

Una vez importados los datos, se renombran campos para poder usarlos en la unión de los datos (figura 6).

Figura 6. Renombrado de campos

```
df_PM2_5.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_O3.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_SO2.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_NO2.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_CO.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)

df_HUM.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_VEL.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_LLU.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_IUV.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_RS.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
df_TMP.rename(columns={'Unnamed: 0': 'fecha'}, inplace=True)
```

Una vez realizado el renombrado de datos, se realiza la unión de los datos usando la fecha y la ciudad como claves de unión dado que se necesita que los datos sean consistentes tanto temporal como espacialmente (figura 7).

Figura 7. Unión de datos

```
df = pd.merge(df_PM2_5, df_O3, on=['fecha', 'ciudad'])
df = pd.merge(df, df_SO2, on=['fecha', 'ciudad'])
df = pd.merge(df, df_NO2, on=['fecha', 'ciudad'])
df = pd.merge(df, df_CO, on=['fecha', 'ciudad'])

df1 = pd.merge(df_HUM, df_LLU, on=['fecha', 'ciudad'])
df1 = pd.merge(df1, df_RS, on=['fecha', 'ciudad'])
df1 = pd.merge(df1, df_VEL, on=['fecha', 'ciudad'])
df1 = pd.merge(df1, df_TMP, on=['fecha', 'ciudad'])
```

Como último paso se reemplaza caracteres en blanco dentro de cada una de las variables por valores perdidos y se eliminan estos datos ya que es necesario contar con un set de datos completos para poder estimar los modelos adecuadamente dando como resultado un set de datos de 634 403 observaciones con 7 variables (figura 8).

Para la unión de los datos atmosféricos

Figura 8. Eliminación de valores perdidos

```
import numpy as np

df.replace('\s+',np.nan, regex=True, inplace=True)
df.dropna(inplace=True)
df['ciudad'] = df['ciudad'].str.capitalize()
print(f'Dimension del set de datos final {df.shape}')

Dimension del set de datos final (634403, 7)
```

Una vez terminadas estas transformaciones se tiene un set de datos con datos para comenzar a realizar el análisis de los resultados.

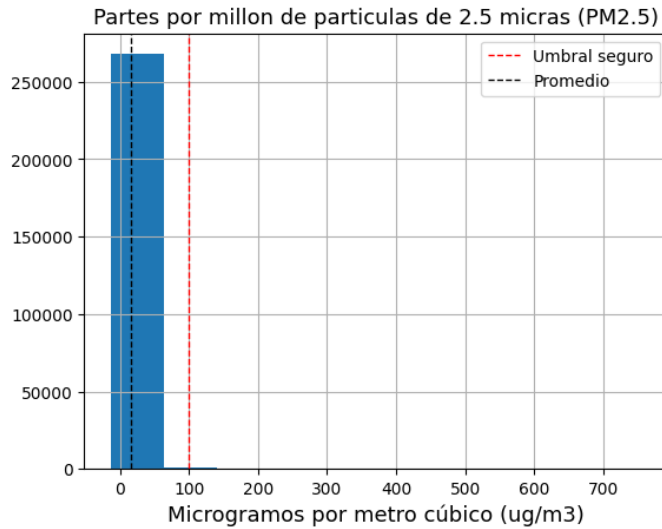
Análisis de resultados

Una vez terminado el proceso de ETL de los datos para el análisis, es necesario dar una visión general del estado en el que se encuentran los datos que ayudaran a predecir la calidad del aire en la ciudad de Quito.

En la figura 9 se observa el comportamiento de las partes por millón de partículas de 2.5 micras en el aire, en este caso el número de partículas en el aire se concentran principalmente por debajo de las 100 partículas por millón, dado que el número por el que al que se comienza a tener problemas pulmonares es de 10 $\mu\text{m}/\text{m}^3$, este número en algunas zonas se puede considerar como comprometedor.

De acuerdo con Linares (2008), el número mínimo al que se relacionan los problemas pulmonares con la calidad del aire es de 10 microgramos por metro cubico ya representa un problema cardiorrespiratorio en la población llegando a provocar cáncer de pulmón.

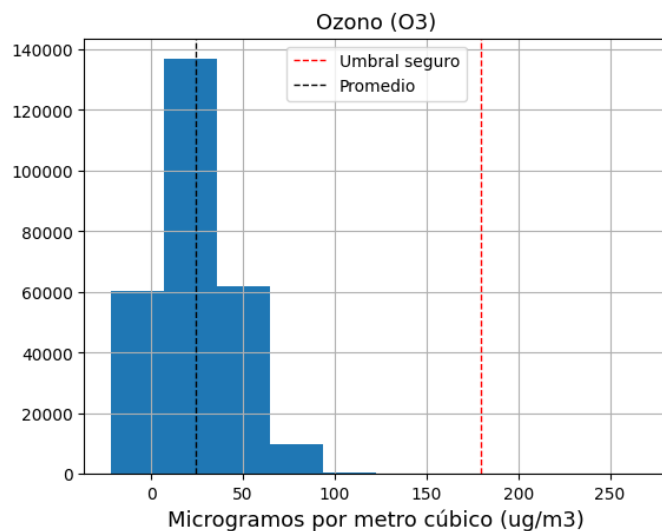
Figura 9. Partes por millón de partículas de 2.5 micras



En el caso del ozono (figura 10) se observa que de forma similar a las partes por millón de partículas de 2.5 micras, la cantidad de partículas de ozono por millón se ubica por debajo de 100 microgramos por metro cuadrado.

En este caso el nivel seguro de ozono se encuentra por debajo de 180 microgramos por metro cúbico, por lo que, de forma general, los niveles de ozono en las zonas de la ciudad de Quito se encuentran en niveles que no afectan en gran medida a la salud de la población.

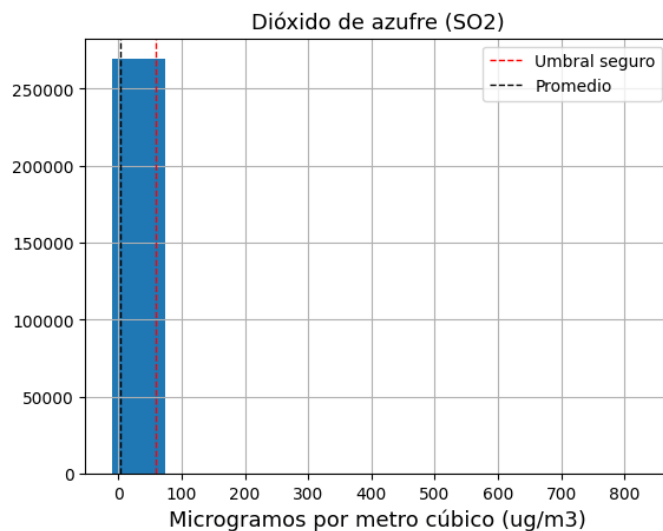
Figura 10. Microgramos de ozono por metro cuadrado



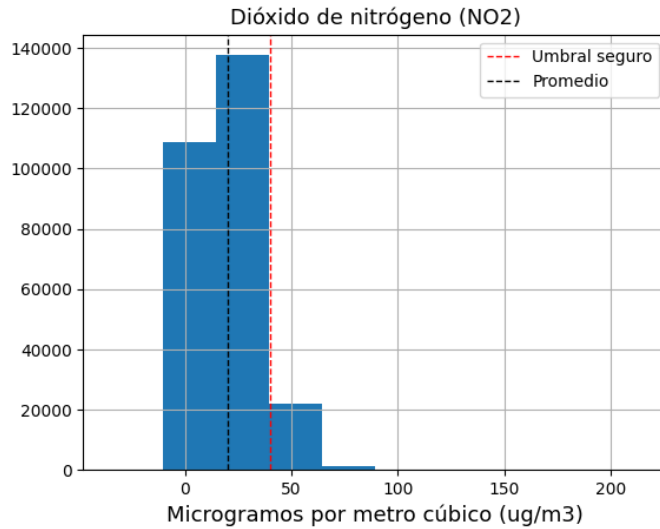
En el caso del dióxido de azufre (figura 11) se observa que en la ciudad de Quito hay un cierto número de zonas y ocasiones en las que tienen un nivel más elevado al considerado como saludable.

En relación a la concentración máxima que no provoca problemas a la salud es de 60 microgramos por metro cúbico, si bien el número de mediciones en las que se sobrepasa el nivel seguro no es muy grande, esto resulta ser una señal de alerta.

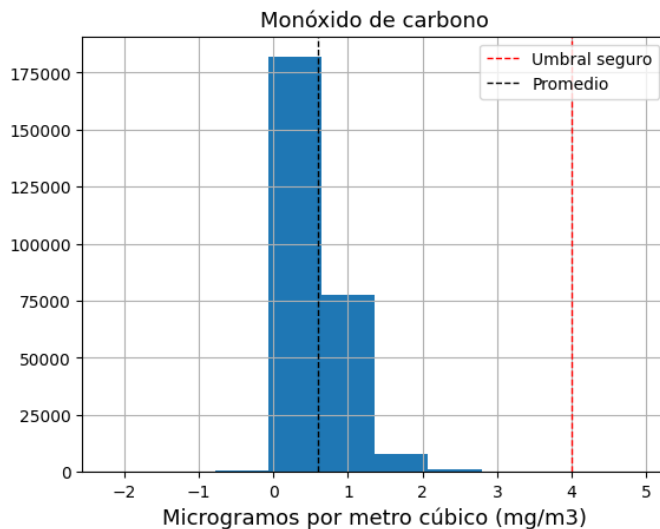
Figura 11. Microgramos de dióxido de azufre por metro cúbico



De la misma forma en el caso del dióxido de nitrógeno (figura 12) se observa que al igual que en el dióxido de azufre, existen zonas y mediciones en las que se sobrepasa los niveles seguros para la salud. A pesar de que no son demasiadas observaciones resulta preocupante dados los principales problemas a la salud que puede provocar en la población.

Figura 12. Microgramos de dióxido de nitrógeno

En cuanto al monóxido de carbono (figura 13) se observa que igual un pequeño número de observaciones sobrepasa el nivel seguro de monóxido de carbono en el aire (4 mg/m³) llegando a 5 miligramos por metro cúbico.

Figura 13. Miligramos de monóxido de carbono por metro cúbico

Si bien los resultados previamente descritos son una forma más amigable de representación de los datos, resulta de utilidad los resultados numéricos, por lo que en la figura 14 se muestran los descriptivos de cada uno de los componentes que se encuentran en el aire y que determinan su calidad.

Haciendo referencia a la cantidad de partículas en el aire de 2.5 micras, en promedio hay 17.04 microgramos por metro cúbico con una desviación estándar de 12.58 microgramos por metro cuadrado que representa una variabilidad amplia dentro de la cantidad de partículas.

Como se mencionó antes, este valor promedio sobrepasa el nivel seguro de 10 microgramos por metro cúbico por lo que representa un aire de menor calidad en la ciudad de Quito que puede provocar problemas de salud en la población de la ciudad.

En el caso del ozono se observa que en promedio existen 24.05 microgramos de ozono por metro cúbico con una desviación estándar de 19.95 microgramos por metro cúbico, en este caso, el nivel de ozono se encuentra dentro de la zona considerada como normal y que no representa un peligro para la salud.

Con respecto a los valores del dióxido de nitrógeno se observa que en promedio se tiene 22.77 microgramos por metro cúbico con una desviación estándar de 13.66 microgramos por metro cúbico, en este caso el promedio de la concentración de dióxido de nitrógeno se encuentra dentro de un rango que no compromete la salud de la población de Quito.

Finalmente, en el caso de la concentración de monóxido de carbono en el aire existe un promedio de 0.73 miligramos por metro cúbico con una desviación estándar de 0.55 miligramos por metro cúbico, se puede observar en la figura 14 que este promedio es inferior al valor más bajo al cual este compuesto afecta la salud respiratoria de una persona.

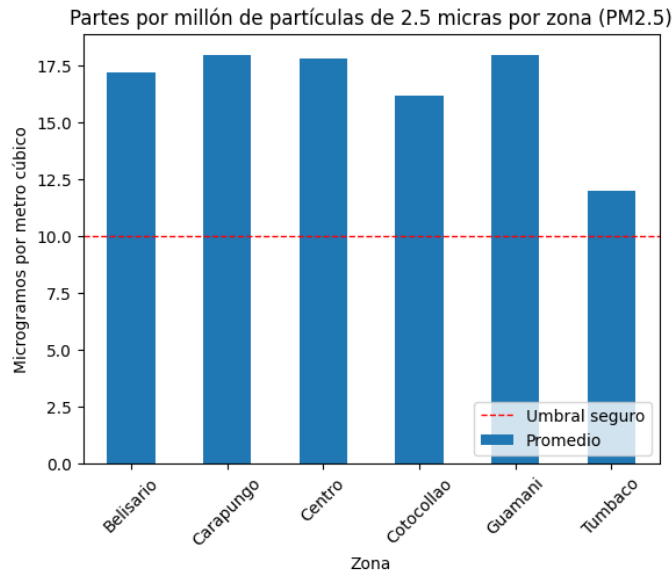
Figura 14. Descriptivos de las características que determinan la calidad del aire

| | PM2_5 | O3 | SO2 | NO2 | CO |
|-------|-------------|-------------|-------------|-------------|-------------|
| count | 634403.0000 | 634403.0000 | 634403.0000 | 634403.0000 | 634403.0000 |
| mean | 17.0423 | 24.0511 | 3.1745 | 22.7718 | 0.7329 |
| std | 12.5829 | 19.9531 | 3.5130 | 13.6637 | 0.5504 |
| min | -82.6700 | -21.8900 | -10.6800 | -35.4500 | -2.2040 |
| 25% | 8.9600 | 6.9400 | 1.3600 | 11.9500 | 0.4080 |
| 50% | 14.8400 | 19.6200 | 2.2900 | 20.9200 | 0.5800 |
| 75% | 22.2200 | 37.1200 | 3.7800 | 31.0900 | 0.8700 |
| max | 837.3400 | 267.5000 | 830.6400 | 182.3400 | 11.9000 |

En un análisis visual de las partes por millón de las partículas de 2.5 micras por zona (figura 15) se observa que no hay ninguna concentración promedio inusual dentro de ciertas zonas, sin

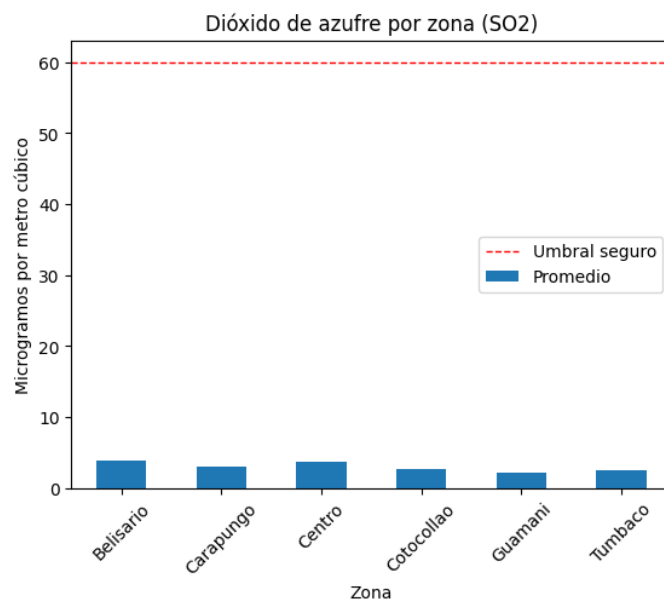
embargo, todas las zonas tienen un promedio más elevado al máximo permitido de 10 microgramos por metro cúbico en el aire, siendo la zona de Tumbaco la que más se aproxima a este nivel ideal.

Figura 15. Partes por millón de partículas de 2.5 micras por zona



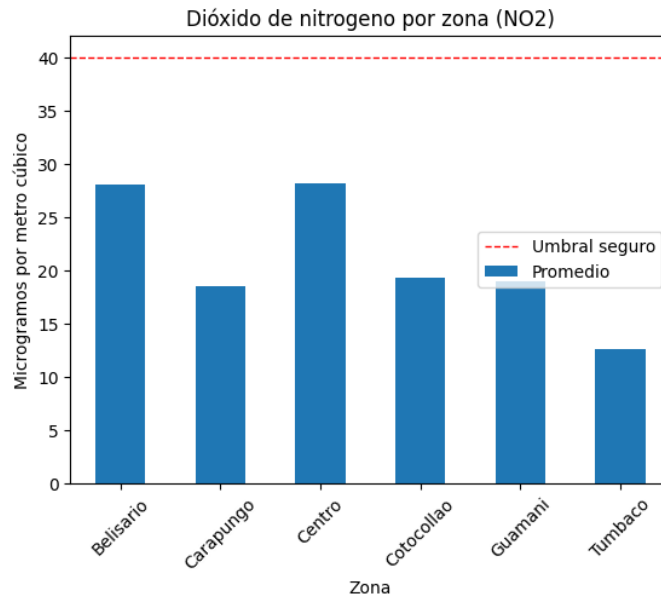
De forma similar en el caso de la concentración de dióxido de azufre por zona (figura 16) se observa que todas las zonas tienen un promedio que se encuentra dentro del rango considerado como seguro, por lo que la concentración de este compuesto en el aire de la ciudad de Quito no representa una complicación para la salud de los quiteños.

Figura 16. Concentración de dióxido de azufre por zona



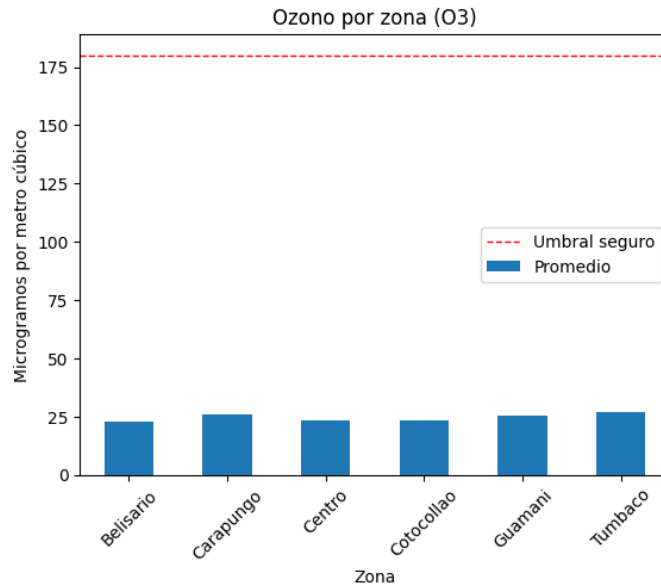
En el caso de la concentración del dióxido de nitrógeno (figura 17) se observa que el nivel de concentración de dióxido de nitrógeno en cada una de las zonas de Quito se encuentra dentro del rango inofensivo para la salud de las personas de Quito.

Figura 17. Concentración de dióxido de nitrógeno por zona



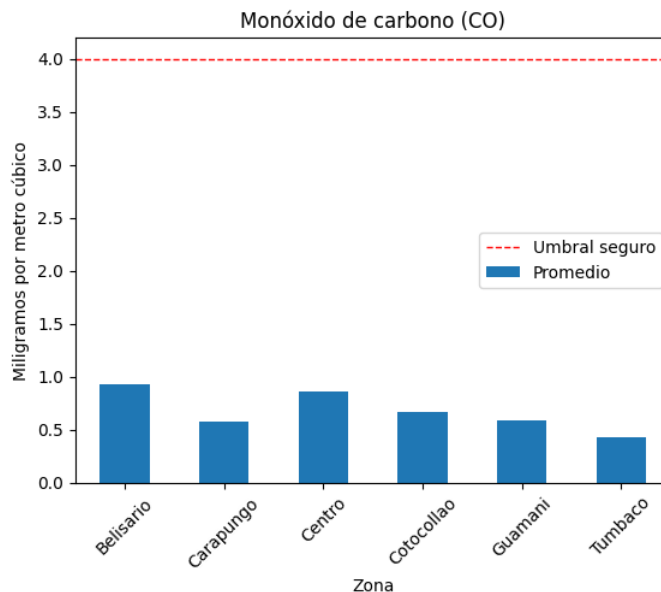
Como se observa en la figura 18, el mismo caso se da en el caso del ozono, se encuentra dentro de un rango en el que se lo puede considerar como inofensivo para la salud de las personas por lo que no se observa ninguna zona cuyo promedio supere el valor establecido como umbral para provocar problemas de salud.

Figura 18. Concentración de ozono por zona



Finalmente, en la concentración de monóxido de carbono por cada una de las zonas en Quito (figura 19) no se observa evidencia de que el promedio de la concentración en cada lugar sobrepase el nivel considerado como seguro para la salud humana desde el 2004 al 2023.

Figura 19. Concentración de monóxido de carbono por zona



Para poder determinar las variables más importantes que influyen en las partes por millón de partículas de 2.5 micras se realiza un análisis de correlación entre las variables, en donde valores cercanos a -1 o 1 representan una correlación fuerte.

Por otro lado, un valor cercado a cero representa una relación débil con la principal variable que determina la calidad del aire, por lo que estas variables no influirían en gran medida a la predicción de la calidad del aire de la ciudad de Quito.

En el análisis de los coeficientes de correlación entre las variables, se observa una correlación media entre el las partes por millón de partículas de 2.5 micras en el aire (P2.5) con la concentración de dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y el monóxido de carbono (CO).

En cuanto al tipo de relación entre las variables, tanto el dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y el monóxido de carbono (CO) tienen una relación positiva con el número de partes por millón de partículas de 2.5 micras.

En el (O₃) se observa que tiene una relación inversa o negativa y débil, esto indica que entre menor sea la concentración de ozono en el aire, mayor será la concentración de partículas en el aire de la ciudad de Quito.

Como se observa en la matriz de correlación (figura 20), en el caso de las variables ambientales como el Dióxido de nitrógeno (NO₂), Dióxido de Azufre (SO₂) y el Monóxido de Carbono (CO) tienen un coeficiente de correlación mayor o igual a 0.3 que son las que más se relacionan.

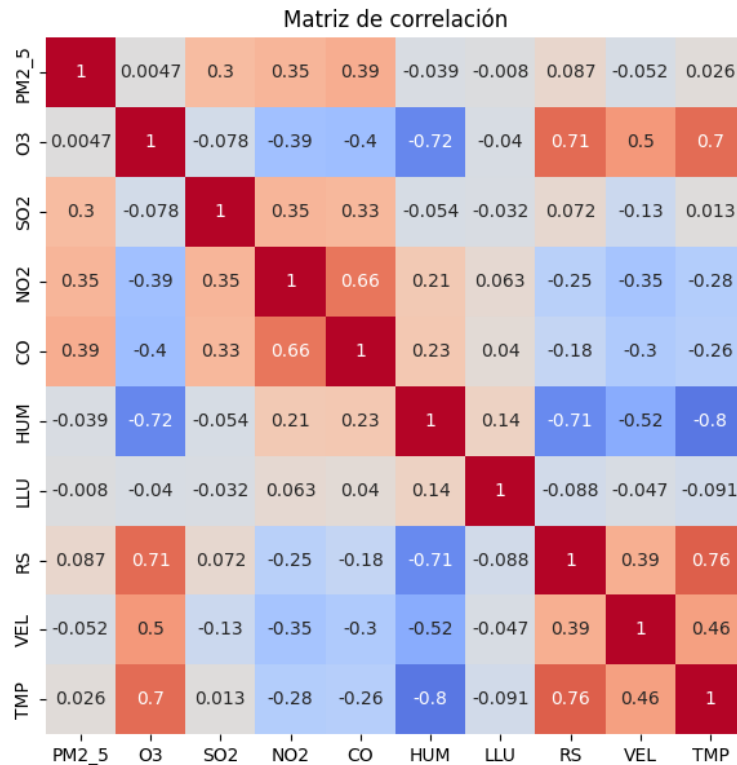
En el análisis de los coeficientes de correlación entre las variables, se observa una correlación media entre el las partes por millón de partículas de 2.5 micras en el aire (P2.5) con la concentración de dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y el monóxido de carbono (CO).

En cuanto a los niveles de Ozono en el aire, a pesar de este factor se toma en cuenta al ser un factor que tiene un efecto sobre la cantidad de partículas de 2.5 micras en el aire.

Por otro lado, en el caso de las variables atmosféricas se observa una correlación casi nula con la cantidad de partículas de 2.5 micras en el aire, esto indica un comportamiento aleatorio de estos factores frente a la calidad del aire.

Además de que no afectan o no se encuentran directamente relacionadas con la calidad del aire ya que empíricamente no hay un efecto meteorológico en la calidad del aire de la ciudad de Quito.

Figura 20. Matriz de correlación



Para determinar el mejor modelo que sea capaz de predecir los valores de la concentración de partículas de 2.5 micras para determinar la calidad del aire en la ciudad de Quito se realizó una experimentación con 5 modelos.

Como se observa en las en la figura 21 la relación entre los compuestos contaminantes, en primer lugar, en el cruce de a relación entre la cantidad de partículas de 2.5 micras y el ozono, en donde se observa la relación débil entre ambas variables dado que la nube de puntos tiene un comportamiento aleatorio.

El dióxido de azufre muestra una relación positiva con el número de partículas de 2.5 micras en el aire, en conjunto con el coeficiente de correlación de 0.3 se puede decir que la relación entre estos datos es positiva, por lo que el aumento en la cantidad de dióxido de azufre, provocara un incremento de partículas en el aire.

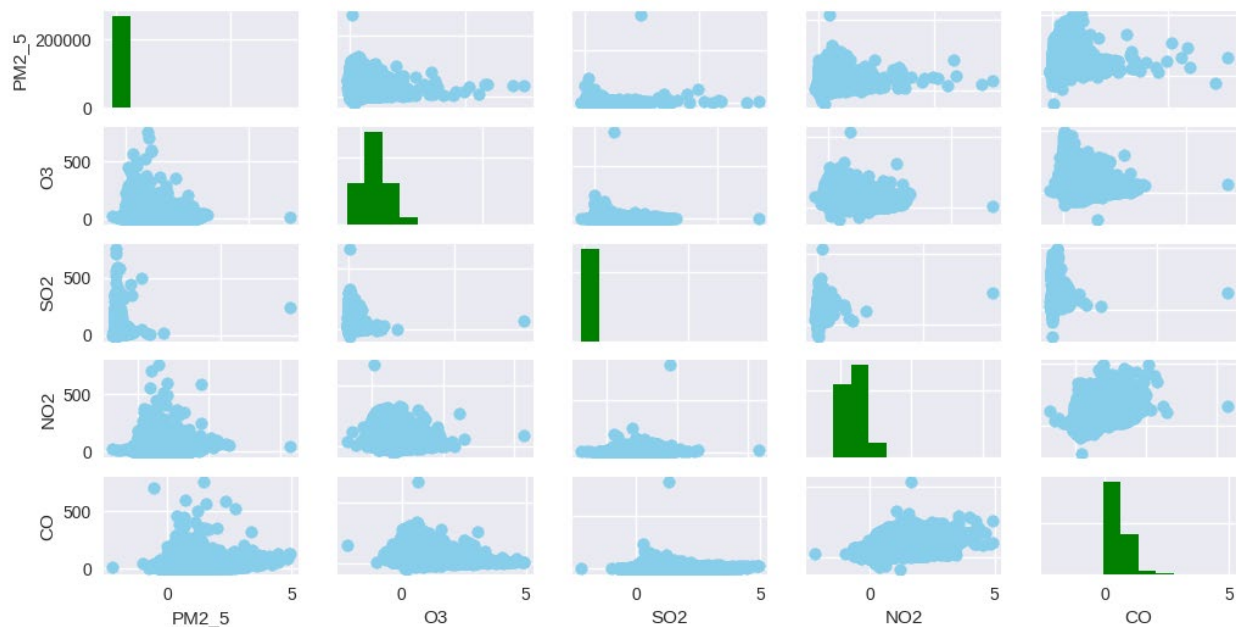
En el caso del dióxido de nitrógeno se observa una tendencia igual al del dióxido de azufre, con un valor de 0.35 se tiene una tendencia creciente, por lo que de igual manera en el caso de que la

cantidad de dióxido de nitrógeno incrementa, también se debería ver un incremento en el caso de las partículas contaminantes en el aire.

Por último, en el caso del monóxido de carbono, se observa un valor de relación un poco más alto (~ 0.40) por lo que la tendencia en el cruce con la variable de la cantidad de partículas contaminantes en el aire es más evidente.

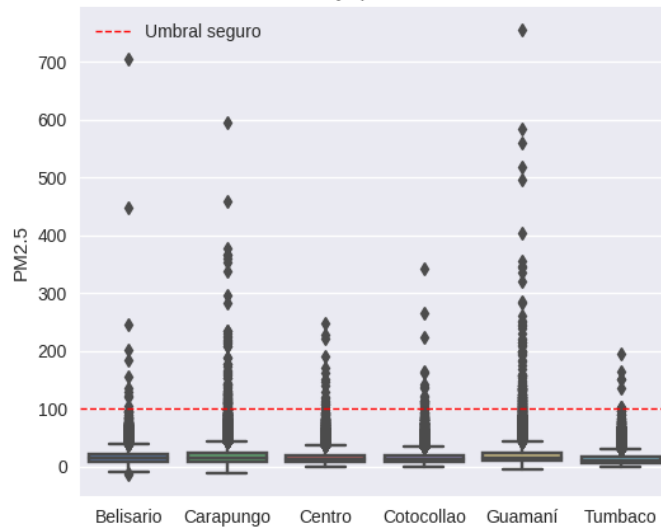
En este caso se muestra una tendencia creciente apoyada con su coeficiente de correlación positiva, por lo que la tendencia de las partículas contaminantes en el aire es creciente con relación al monóxido de carbono.

Figura 21. Matriz de dispersión de compuestos contaminantes



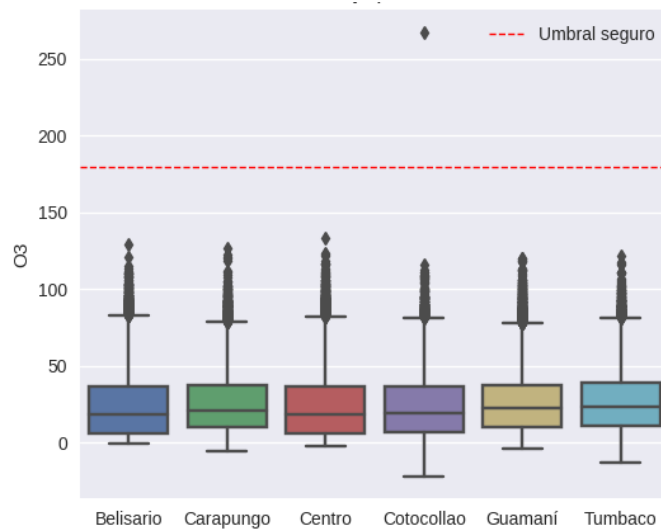
En cuanto a las variables más significativas dentro del análisis, se puede visualizar en la figura 22 que en todas las zonas de la ciudad de Quito normalmente no se supera un nivel seguro en la cantidad de partículas contaminantes en el aire, sin embargo, existen ciertas mediciones que sobrepasan este nivel seguro de la concentración de partículas.

Figura 22. Gráfico de caja de las partes por millón de partículas de 2.5 micras por zona



Por otro lado, se observa en el caso del ozono que no se encuentran mediciones normales y consideradas como atípicas en cada una de las zonas, en el caso de la zona de Cotocollao se ve un valor atípico superior a los 250 microgramos por metro cubico que es muy superior al considerado como seguro para la salud de la población de la ciudad de Quito.

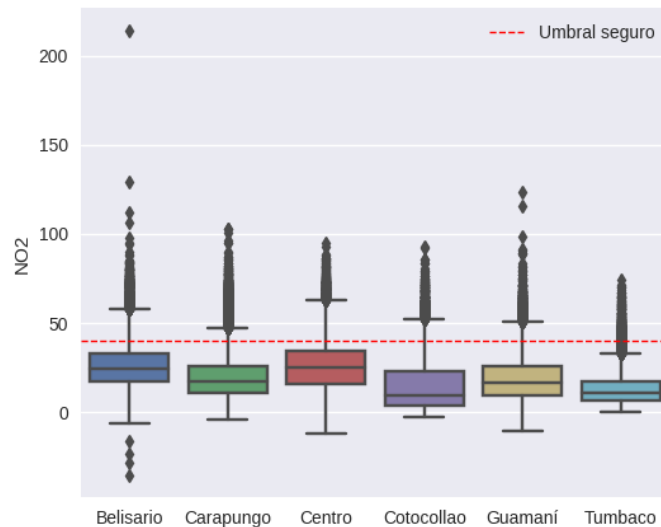
Figura 23. Gráfico de la concentración de ozono por zona



En cuanto a la figura 25 se observa la concentración de dióxido de nitrógeno por zona en donde se observa que existen mediciones que sobrepasan los niveles seguros para la salud de la población.

Este aspecto resulta preocupante dado que incluso los valores que se consideran como mediciones normales dentro del gráfico de caja se encuentran por encima de los niveles dañinos para la salud de la población de estas zonas de la ciudad de Quito.

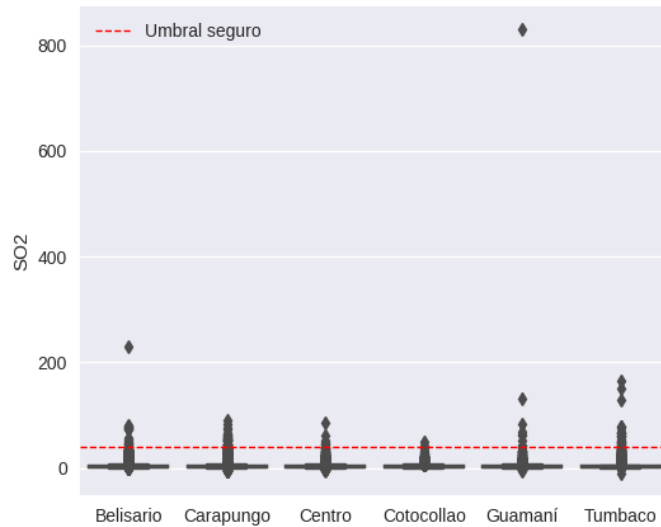
Figura 24. Concentración de dióxido de nitrógeno por zona



En el caso del dióxido de azufre de igual manera se observan valores atípicos que sobrepasan por mucho el umbral seguro, esto refleja una posible mala calidad del aire en la ciudad de Quito.

A pesar de esto, resulta estar en una mejor situación que en el caso de dióxido de nitrógeno dado que las mediciones que pasan el umbral seguro son concentraciones o mediciones raras del dióxido de azufre.

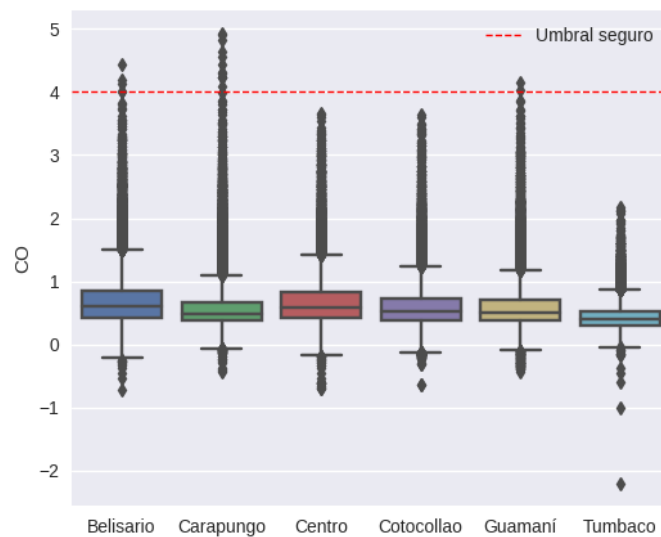
Figura 25. Concentración del dióxido de azufre por zona



Por último, en el caso del monóxido de carbono se observa un resultado similar al de la concentración de azufre, únicamente son concentraciones poco comunes de carbono en microgramos por metro cubico se encuentran por encima de un valor seguro o inofensivo para la salud humana.

Sin embargo, es necesario controlar estas mediciones y disminuir estos niveles ya que la calidad del aire se ve comprometida, ergo, la salud de la población.

Figura 26. Concentración de monóxido de carbono por zona



Experimentación

Los mencionados anteriormente corresponden a un modelo de regresión lineal múltiple, una regresión mediante vectores de soporte (SVR), regresión mediante árbol de predicción, regresión mediante vecinos más cercanos (KNN) y por último un modelo de red neuronal para regresión.

A continuación, en la tabla 1 se presenta las configuraciones para la estimación de los modelos de regresión mediante cada uno de los algoritmos de aprendizaje automático.

Tabla 1. Configuración de hiperparámetros

| Modelo | Configuración de hiperparámetros |
|----------------------------|---|
| Regresión lineal | Función de optimización: Mean Squared Error (MSE) |
| Modelo de SVR | Kernel: Radial Basis Function (RBF) |
| Árbol de predicción | Función de optimización: Mean Squared Error (MSE) |
| Vecinos más cercanos (KNN) | Número de vecinos: 5 Distancia: Minkowski |

En el caso de la red neuronal se presenta la estructura de la red neuronal (figura 21), en el caso de los demás hiperparámetros se tiene a la función Mean Squared Error (MSE) como función a minimizar, un optimizador Adam, entrenado por 20 épocas y con un tamaño de batch de 128.

En cuanto a la topología de la red neuronal se observa que se tiene una capa de entrada (InputLayer) con un input de 4 dado que son 4 variables independientes con las que el modelo será entrenado.

La primera capa oculta del modelo (hidden_layer_1) cuenta con un total de 64 neuronas con una función de activación ReLU, en cuanto a la segunda capa oculta (hidden_layer_2), únicamente tiene 32 neuronas con una función de activación ReLU y por último la capa de salida (output_layer) tiene únicamente una neurona con una función de activación lineal dado que el problema afrontado en el presente trabajo es de regresión.

Esta configuración de hiperparámetros da como resultado un total de 2433 parámetros entrañables dentro de la red neuronal.

Figura 27. Topología de la red neuronal

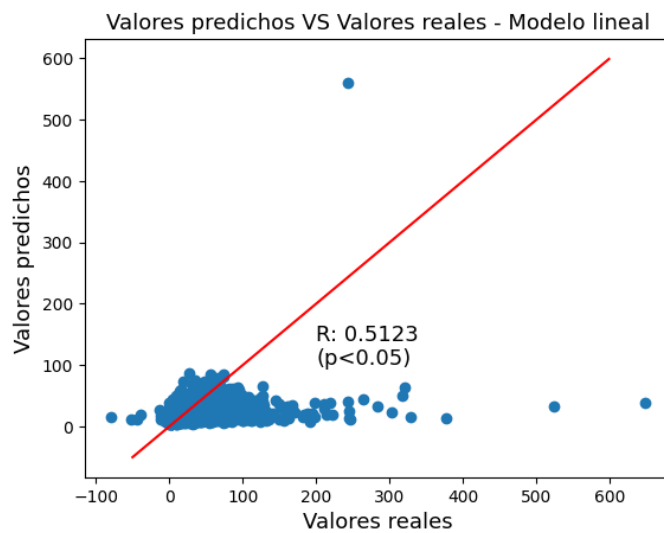
```

Model: "ann_model"
=====
Layer (type)              Output Shape              Param #
-----
input_3 (InputLayer)      [(None, 4)]              0
hidden_layer_1 (Dense)    (None, 64)              320
hidden_layer_2 (Dense)    (None, 32)              2080
output_layer (Dense)      (None, 1)               33
=====
Total params: 2433 (9.50 KB)
Trainable params: 2433 (9.50 KB)
Non-trainable params: 0 (0.00 Byte)
    
```

Una vez realizada la estimación de los resultados del modelo, es necesario comprobar si es que los resultados obtenidos mediante el modelo de predicción se apegan lo suficiente a los valores reales para poder determinar la validez de los resultados.

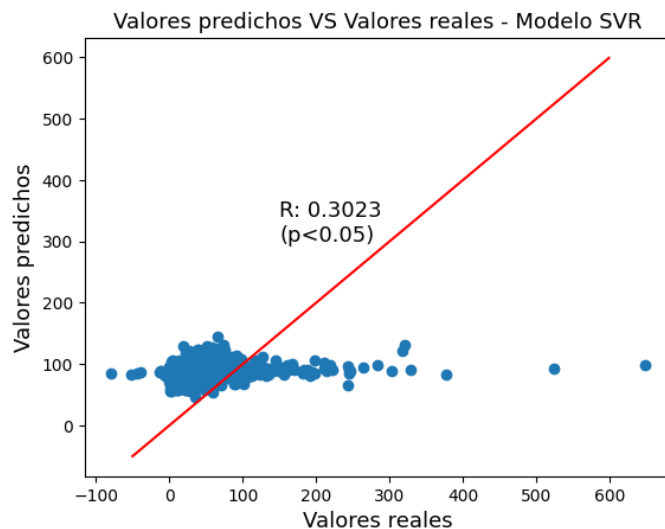
En el caso del modelo lineal (figura 29), se observa que el coeficiente de correlación entre los valores predichos y los valores reales de los datos tienen una correlación media y positiva, lo que indica que los valores predichos siguen la misma tendencia que los datos reales, además de que esta relación es estadísticamente significativa ($p < 0.05$).

Figura 28. Valores predichos VS valores reales – Modelo lineal

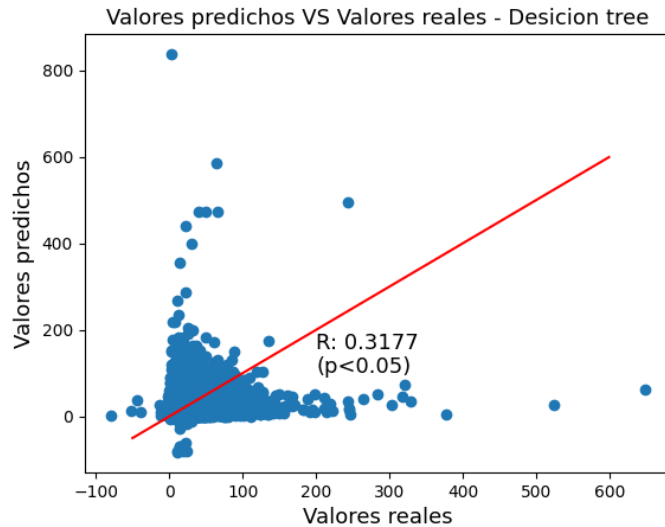


De la misma forma, se observa que en el caso de los valores predichos por el modelo de regresión usando vectores de soporte (figura 30) y los valores reales tienen una correlación positiva y estadísticamente significativa ($p < 0.05$) lo que refleja que este modelo cumple con el cometido de realizar predicciones válidas.

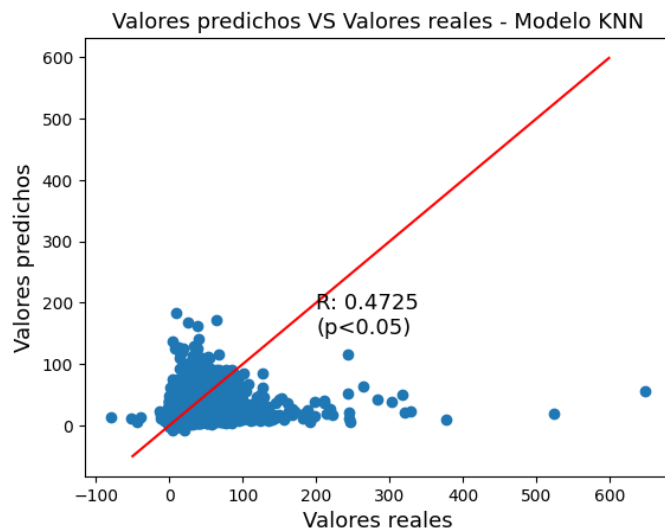
Figura 29. Valores predichos VS valores reales – Modelo SVR



Igualmente, en el caso de la regresión usando el árbol de predicción (figura 31) se obtiene una correlación positiva y estadísticamente significativa entre los valores predichos y los valores reales de la concentración de partículas de 2.5 micras en el aire de Quito, esto determina que los valores predichos siguen la misma tendencia de los valores reales de la calidad del aire de la ciudad de Quito.

Figura 30. Valores predichos VS valores reales – Modelo de árbol de predicción

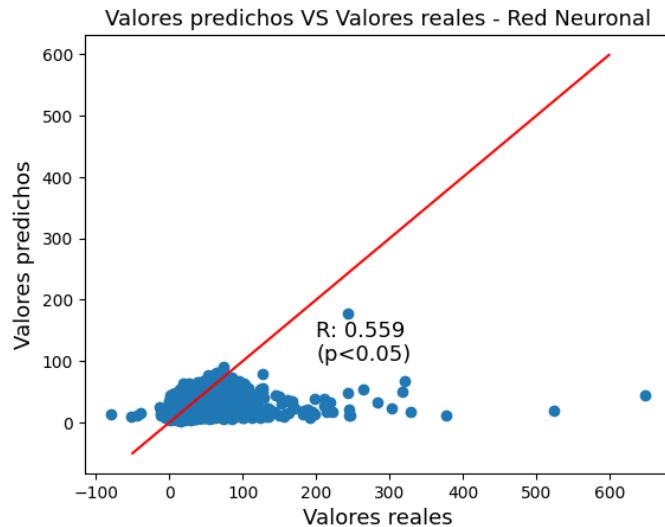
En el resultado del modelo de regresión usando vecinos más cercanos (figura 32) se observa que la relación entre los valores reales y predichos de la concentración de partículas de 2.5 micras en el aire de Quito tienen una relación media y positiva lo que refleja una predicción adecuada por parte del modelo de regresión, además de que esta relación es estadísticamente significativa ($p < 0.05$).

Figura 31. Valores predichos VS valores reales – Modelo KNN

Finalmente se observa la misma tendencia que anterior en el modelo estimado usando la red neuronal (figura 33) se observa una relación media y positiva entre valores predichos y valores

reales de la concentración de partículas de 2.5 micras en el aire de Quito, además de que la relación es estadísticamente significativa ($p < 0.05$).

Figura 32. Valores predichos VS valores reales – Modelo de red neuronal



Como se ha visto previamente, todos los modelos de predicción son capaces de aproximar los valores de la calidad del aire de la ciudad de Quito a partir de la cantidad de partículas de 2.5 micras en el aire, sin embargo, es necesario determinar cuál es el mejor modelo de predicción para esta tarea.

Para dar una respuesta objetiva a este objetivo se presentan las métricas de comparación de cada uno de los modelos.

En la tabla 2 se observa que, de acuerdo con las métricas, el mejor modelo de predicción resulta ser el que es basado en redes neuronales ya que cuenta con los errores más consistentes de todos los demás modelos de predicción tanto en MSE, MAE y RMSE, además de que cuenta con el R2 más alto en entrenamiento que comparado con el R2 en el entrenamiento no se encuentran muy alejados.

En el caso del modelo lineal se obtiene un modelo similar al de las redes neuronales ya que cuenta con un R2 cercano tanto en el entrenamiento como en la prueba del modelo, esto refleja que este modelo también resulta consistente y semejante al de las redes neuronales, por lo que es una alternativa viable en el caso de usar un modelo alternativo para poder predecir la calidad del aire en la ciudad de Quito.

En el caso de los otros modelos de aprendizaje no se puede constatar que sean apropiados para predicción ya que cuentan con una capacidad explicativa muy baja en las pruebas, pero en el set de entrenamiento sobrepasan el 40%, esto es un indicio de sobre ajuste, lo que no permite considerar a los modelos como apropiados para una buena predicción.

Tabla 2. Métricas de evaluación de los modelos de predicción

| Modelo | Training | | | | Test | | | |
|---------------------|----------|-----------|---------|---------|--------|-----------|---------|---------|
| | R2 | MSE | MAE | RMSE | R2 | MSE | MAE | RMSE |
| Lineal | 0.2365 | 108.9693 | 6.7091 | 10.4388 | 0.2166 | 122.8624 | 6.7600 | 11.0800 |
| Vectores de soporte | 0.1233 | 3675.5075 | 59.7329 | 60.6260 | 0.1087 | 3692.1929 | 59.8054 | 60.7634 |
| Árbol de predicción | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0594 | 237.6004 | 9.5389 | 15.4143 |
| KNN | 0.4160 | 83.4354 | 5.9304 | 9.1343 | 0.1494 | 139.5071 | 7.3004 | 11.8113 |
| Red neuronal | 0.2201 | 128.1220 | 6.6562 | 11.3191 | 0.2433 | 113.818 | 6.6245 | 10.6685 |

Análisis con el mejor modelo estimado

De acuerdo con el resultado previo, se observó que el mejor modelo resultó ser los del modelo lineal y la red neuronal, de estos modelos, a continuación, se presentan las métricas de evaluación con los datos no vistos aun por el modelo entrenado.

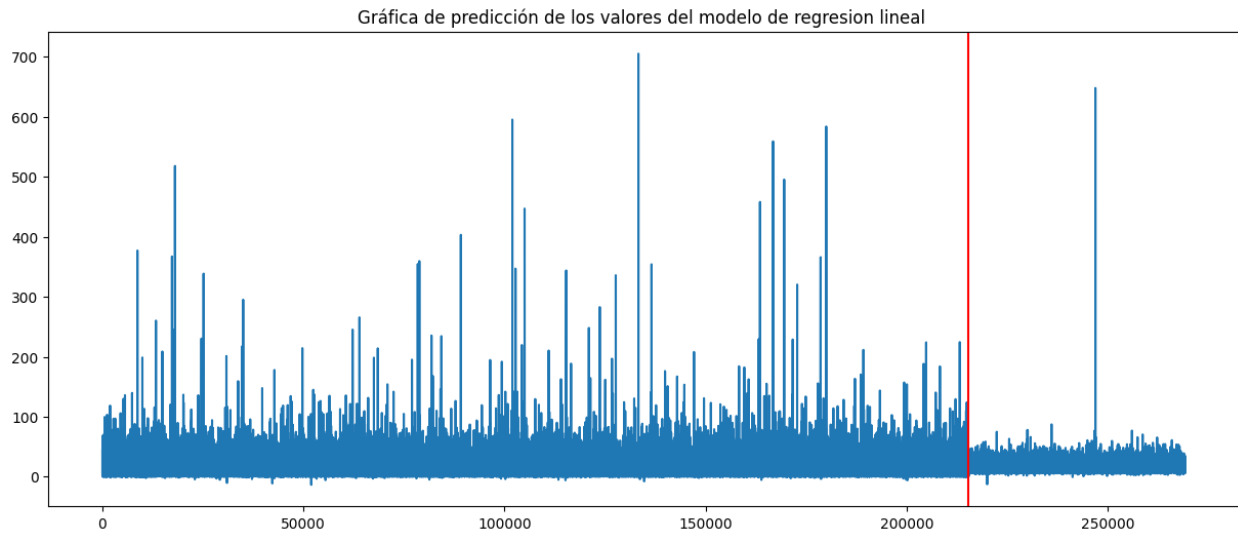
| | Modelo Lineal | Red Neuronal |
|------|---------------|--------------|
| MSE | 122.8624 | 113.818 |
| MAPE | 6.76 | 6.6245 |
| RMS | 121.6843 | 112.6105 |
| MAX | 168.4265 | 112.6403 |

Para poder expresar y visualizar de mejor manera los resultados obtenidos con los dos mejores modelos, se observa que la predicción de la contaminación en la ciudad de Quito es consistente con la demás información de la serie temporal.

Como se observa en la Figura 33, el pronóstico de la serie temporal sigue la misma tendencia estacionaria que la serie de datos original, esto implica que las predicciones realizadas con este modelo serán consistentes en el tiempo.

Visualmente, se puede observar que los valores predichos tienen fluctuaciones similares a los datos previos, reflejando aproximaciones y predicciones robustas a lo largo de un tiempo considerable.

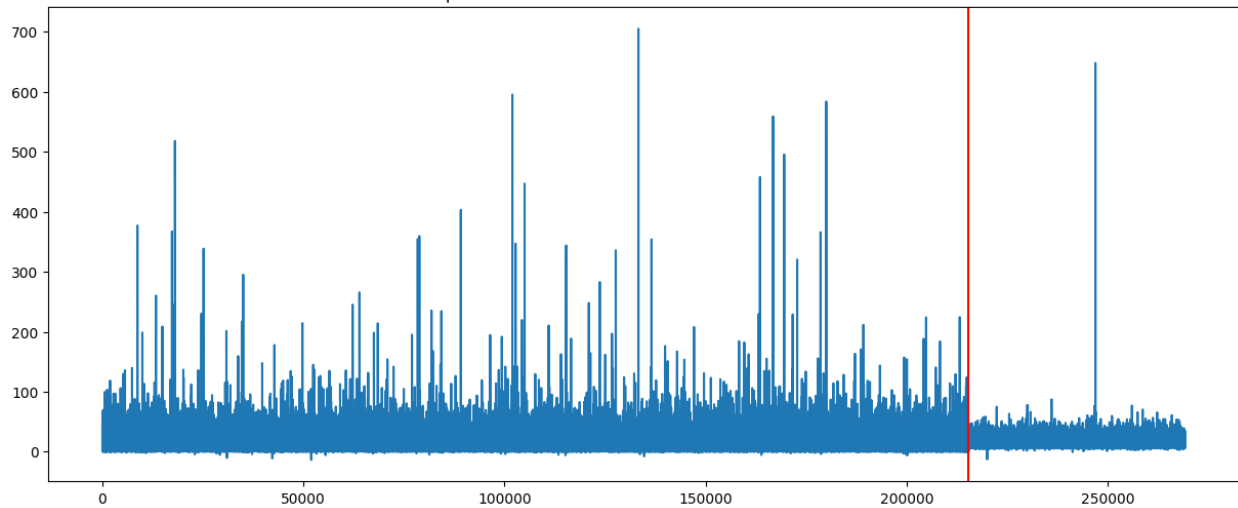
Figura 33. Pronostico temporal usando regresión lineal



En el caso del modelo con redes neuronales, se observa un resultado similar al de los resultados de la regresión lineal, este resultado es considerado como el más esperado a que ambos modelos tuvieron métricas similares durante la evaluación de los mismos.

Del mismo modo, se obtiene un resultado de pronósticos similares a los datos usados para el entrenamiento, lo que también respalda la robustez de la capacidad predictiva del modelo para poder estimar los valores de la calidad del aire dentro de la ciudad de Quito, en una comparativa se observa un comportamiento similar en ambas tendencias.

Gráfica de predicción de los valores del modelo de redes neuronales



Análisis temporal

Dentro del análisis temporal, se analizan 4 componentes principales que componen un análisis de tiempo de una variable, dentro de estas se encuentra la serie observada, la tendencia, la estacionariedad y los residuos.

Tendencia: La tendencia representa la dirección general de la serie a lo largo del tiempo. Indica si la serie está aumentando, disminuyendo o permanece relativamente constante. La tendencia puede ser lineal o no lineal.

Estacionalidad: La estacionalidad refleja patrones periódicos o cíclicos que se repiten a intervalos regulares. Pueden ser patrones diarios, semanales, mensuales, anuales u otros. La estacionalidad muestra cómo la serie varía de manera sistemática en diferentes momentos del año o del período.

Ciclo: El ciclo es una fluctuación recurrente, pero no necesariamente periódica, que no está asociada con patrones estacionales fijos. A menudo, los ciclos tienen duraciones más largas y pueden estar relacionados con factores económicos, sociales u otros fenómenos.

Residuos: Los residuos representan la variación no explicada por la tendencia, la estacionalidad y el ciclo. Son la parte aleatoria o "ruido" de la serie temporal. Si la descomposición es efectiva, los residuos deberían parecer ruido blanco, es decir, no deberían mostrar patrones discernibles.

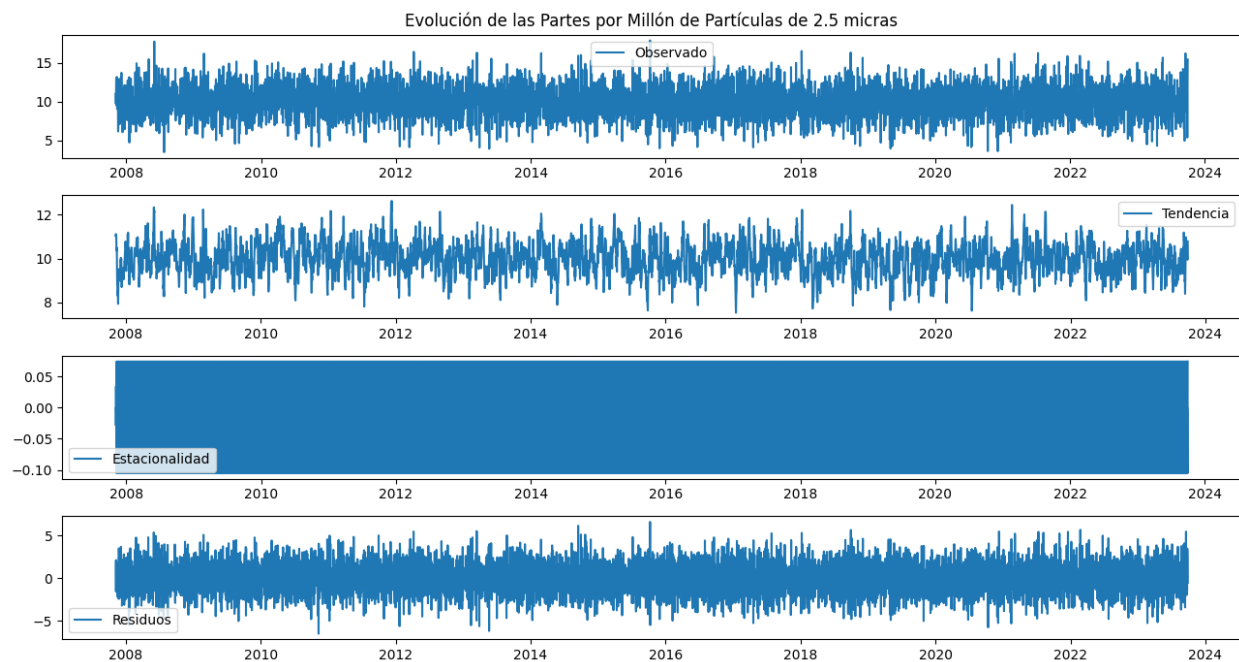
Según se observa en el resultado de la descomposición de la serie temporal, en el caso de la tendencia observada, se observa una tendencia estacionaria dentro de la serie temporal, lo que

refleja que el nivel de contaminación aparentemente se encuentra estable, por lo que no hay un aumento o decaimiento de las partículas contaminantes en la ciudad de Quito.

En el caso de la tendencia, se recalca el resultado obtenido previamente, en el cual no se observa una tendencia de subida o bajada dentro de la serie temporal, por lo que su comportamiento es constante y estable a lo largo del tiempo desde enero del 2008 hasta finales de 2023.

En cuanto al componente estacional se observa una tendencia cíclica que se repite dentro de todo el periodo de análisis, esto refleja escenarios similares a lo largo del tiempo que son comportamientos normales dentro de una serie longitudinal, dado que existen costumbres o comportamientos que se mantienen y se manifiestan en situaciones concretas a lo largo del tiempo.

Por último, en el caso del análisis de los residuales, se observa que su comportamiento sigue un ruido blanco¹ por lo que se puede decir que el comportamiento de la contaminación en la ciudad de Quito es constante y las predicciones realizadas con esta información resulta de calidad.



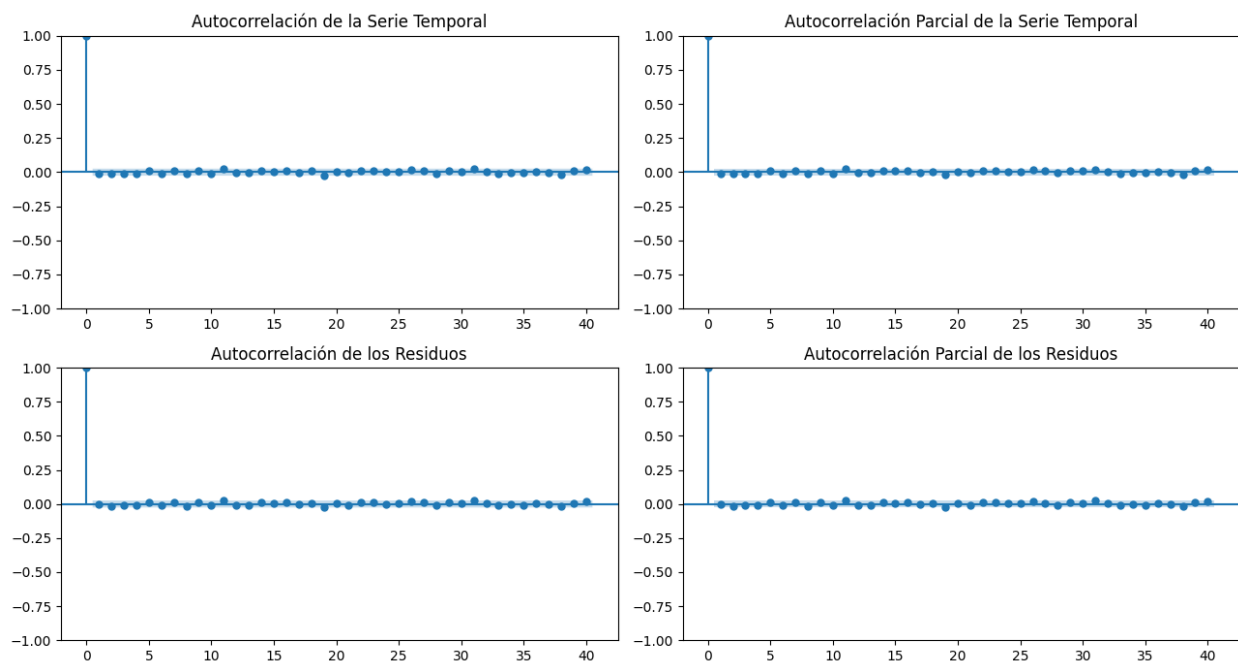
¹ El ruido blanco es un comportamiento característico de las series temporales que indica un comportamiento constante de su variabilidad además de que se encuentra centrado en cero.

Para poder determinar el modelo que se va a utilizar se realiza un análisis mediante los correlogramas los cuales son representaciones gráficas que ofrecen información sobre la autocorrelación en una serie temporal, es decir, la relación de la serie consigo misma en diferentes intervalos de tiempo.

La función de autocorrelación (ACF) muestra la correlación directa entre los valores de la serie en distintos rezagos, mientras que la función de autocorrelación parcial (PACF) indica la correlación condicional eliminando la influencia de los rezagos intermedios.

Los picos y caídas en estas funciones revelan patrones de dependencia temporal, como estacionalidades, ciclos o tendencias, siendo esenciales para seleccionar modelos adecuados de series temporales, como los modelos ARIMA, y diagnosticar la presencia de autocorrelación en los residuos del modelo.

En este caso, se observa que no hay una correlación fuerte entre los datos respecto a la misma tendencia con otros rezagos, de la misma forma en el caso de la correlación parcial, no hay una correlación fuerte entre los demás rezagos de la serie temporal, por lo que se indicaría que el comportamiento de la contaminación en Quitopar más que ser variable, tiene un comportamiento aleatorio.



Utilizando Python mediante un modelo SARIMAX que evalúa tanto modelos AR, MA o su combinación ARIMA para poder lograr el mejor resultado, se observa que dentro del modelo se llegó a la conclusión que el mejor modelo se constituye como un modelo ARIMA(1,1,1) lo que implica que el mejor modelo tiene un componente autoregresivo, un componente de medias móviles y un componente de diferenciación primera diferencia de la serie temporal.

```

SARIMAX Results
=====
Dep. Variable:          ppm_2.5      No. Observations:      5809
Model:                 SARIMAX(1, 1, 1)x(0, 1, 1, 12)  Log Likelihood         -12257.378
Date:                  Wed, 17 Jan 2024      AIC                    24522.756
Time:                  00:43:05          BIC                    24549.415
Sample:                11-05-2007          HQIC                   24532.030
                    - 09-30-2023
Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ar.L1         -0.0152      0.013       -1.154     0.248     -0.041     0.011
ma.L1         -0.9989      0.001     -700.100     0.000     -1.002     -0.996
ma.S.L12      -0.9989      0.007    -150.427     0.000     -1.012     -0.986
sigma2         3.9648      0.075     52.530     0.000      3.817     4.113
=====
Ljung-Box (L1) (Q):      0.00  Jarque-Bera (JB):      1.11
Prob(Q):                 0.97  Prob(JB):              0.57
Heteroskedasticity (H):  1.01  Skew:                  0.01
Prob(H) (two-sided):    0.75  Kurtosis:              3.06
=====

```

En cuanto a las predicciones realizadas por el modelo de series de tiempo, se observa un error pequeño dentro de las predicciones las cuales son capaces de realizarse de forma confiable hasta mediados o comienzos del año 2024.

Conclusiones

Durante el proceso de investigación y análisis realizado en este trabajo se ha logrado obtener información útil y objetiva acerca de aspectos relacionados ambientales de la ciudad de Quito, a continuación, se detallan algunas de las principales conclusiones a las que se ha logrado llevar durante la realización del presente trabajo de investigación.

La única característica que se puede destacar que resulta un posible riesgo para la salud respiratoria de los habitantes de la ciudad de Quito es el de la concentración de partículas de 2.5 micras.

Con respecto a los demás compuestos se observa que de momento se encuentran dentro de un rango considerado como inofensivo a pesar de que existen ciertas observaciones que sobrepasan sus respectivos umbrales de tolerancia.

En un análisis individual del promedio de partículas de 2.5 micras promedio en el aire de varias zonas de la ciudad de Quito se observó que todas las zonas tienen un nivel promedio muy superior al considerado como seguro, por lo que esto puede repercutir en la salud de los habitantes. Esto resulta preocupante ya que esta característica puede desembocar en problemas pulmonares y cardiorrespiratorios graves como el cáncer de pulmón.

En el caso de aspectos influyentes para la predicción de los modelos solo se toma en cuenta características relacionadas a compuestos contaminantes directamente relacionados con la calidad del aire como el dióxido de azufre (SO₂), monóxido de carbono (CO), dióxido de nitrógeno (NO₂) y el ozono (O₃) por su correlación estadísticamente significativa con la cantidad de partículas de 2.5 micras en el aire ($p < 0.05$).

Por otro lado dado que los aspectos relacionados con las variables meteorológicas cuentan con una relación con la calidad de aire casi nula ($R \sim 0$) no se puede considerarlos como candidatos válidos de predicción de la calidad del aire en la ciudad de Quito.

Dando respuesta al principal objetivo del presente trabajo de investigación se puede concluir que el modelo más adecuado para predecir la calidad del aire de la ciudad de Quito mediante la cantidad de partículas de 2.5 micras es el de redes neuronales ya que son las que cometen el menor error de todos los demás modelos.

Asimismo, el modelo de regresión lineal múltiple es capaz de obtener resultados similares al de las redes neuronales, por lo que la regresión lineal es una alternativa muy viable en caso de que se desee tener un modelo más interpretable o intuitivo (modelo de caja blanca) ya que el modelo de red neuronal no resulta tan explicativo a simple vista (modelo de caja negra).

Para poder mitigar este tipo de factores contaminantes es necesario hallar alternativas renovables de energía, recursos y cambiar el paradigma de un proceso lineal por uno cíclico, en donde los materiales se puedan reusar y mitigar los factores contaminantes resultado de los procesos de fabricación de materiales aún más contaminantes.

Al examinar la evolución temporal de estas partículas, es posible identificar patrones estacionales, tendencias a largo plazo, y otros comportamientos que pueden tener implicaciones significativas para la salud pública y el medio ambiente.

El uso de herramientas como la descomposición de series temporales, los correlogramas y modelos específicos, como los modelos SARIMA, permite capturar y modelar la complejidad inherente a estas series.

Trabajos futuros

Dentro de la línea de trabajos futuros hay algunas consideraciones usando como base los avances y resultados del presente trabajo de investigación. Para ello se plantean las siguientes propuestas como línea de trabajo futuro para poder ampliar la usabilidad de temas relacionados con el presente trabajo con el fin de mejorar la calidad de vida de la población.

Como primer punto se propone aplicar los modelos estimados dentro del presente trabajo de investigación para generar herramientas de predicción en tiempo real conectado a una aplicación web para que los habitantes puedan mantenerse al tanto de la calidad de medio ambiente.

Como segundo punto se propone aplicar los modelos estimados para poder anticipar posibles problemas o avances en el daño medioambiental para generar políticas desde el municipio que permitan mitigar estos efectos.

Como último punto se plantea una idea de extrapolación del modelo a nivel nacional, es decir, realizar un análisis similar a nivel país para mejorar la calidad de vida no solo de una ciudad sino de todas las personas que se encuentran dentro del territorio nacional.

Bibliografía

- Aránguez, E. (2019). *CONTAMINANTES ATMOSFÉRICOS Y SU VIGILANCIA*. Obtenido de https://scielo.isciii.es/pdf/resp/v73n2/contam_atmos.pdf
- Chicón, L. (2023). *OZONO ATMOSFÉRICO*. Obtenido de https://portalacademico.cch.unam.mx/materiales/prof/matdidac/sitpro/exp/quim/quim1/quimI_vall/ozono.pdf
- Chiluisa, M. (2023). *Tipos de contaminación y sus principales consecuencias*. Obtenido de <https://www.fundacionaquae.org/wiki/tipos-contaminacion/>
- Comunidad de Madrid,. (2021). *OZONO (O3) TROPOSFÉRICO*. Obtenido de https://www.comunidad.madrid/sites/default/files/doc/sanidad/samb/ficha_ozono_profesionales_2021.pdf
- Gallegos, F. (2018). *DIÓXIDO DE NITRÓGENO*. Obtenido de <https://www.insst.es/documents/94886/431980/DLEP+116++Di%C3%B3xido+de+nitr%C3%B3geno++A%C3%B1o+2018.pdf/b295ec76-11fc-43fb-9834-31e8336d99bc?version=1.0&t=1551310405104>
- Ibarlucía, D. (2019). *Determinación del NO2 (Dióxido de Nitrógeno) atmosférico en la ciudad de Tandil*. Obtenido de <https://ridaa.unicen.edu.ar:8443/server/api/core/bitstreams/8ac1af56-79ac-4db1-b9e8-f43d27d5779a/content>
- Malagón, E. (2011). *Medio Ambiente y Contaminación. Principios básicos*. Obtenido de Attribution-NonCommercial-NoDerivatives 4.0 International: <https://addi.ehu.es/bitstream/handle/10810/16784/Medio%20Ambiente%20y%20Contaminaci%C3%B3n.%20Principios%20b%C3%A1sicos.pdf>
- Narváez, R. P. (2021). *Estrategia para el desarrollo de un modelo de predicción de la calidad del aire de Quito y resultados preliminares de la fase de diagnóstico*. Obtenido de http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1683-07892006000100008
- Organización Mundial de la Salud. (2019). *Dióxido de Nitrógeno*. Obtenido de <https://www.greenfacts.org/es/dioxido-nitrogeno-no2/dioxido-nitrogeno-no2-greenfacts.pdf>

- Pallo, S. (2021). *La contaminación atmosférica, el aire que respiramos*. Obtenido de https://www.aragon.es/documents/20127/674325/CONTAMINACION_ATMOSFERIC_A.pdf/9271acea-8a9b-de61-900f-2ae65565a69e
- Paredes, C. (2020). *Estrategia para el desarrollo de un modelo de predicción de la calidad del aire de Quito y resultados preliminares de la fase de diagnóstico*. Obtenido de http://www.scielo.org.bo/scielo.php?pid=S1683-07892006000100008&script=sci_abstract
- Placeres, M. R. (2020). *La contaminación del aire: su repercusión como problema de salud*. Obtenido de <https://www.redalyc.org/pdf/2232/223214848008.pdf>
- Rodríguez, M. C. (2017). *Las Partículas Respirables PM10 y su Composición Química en la Zona Urbana y Rural de Mexicali, Baja California en México*. Obtenido de <https://scielo.conicyt.cl/pdf/infotec/v25n6/art03.pdf>
- Sango, G. (2018). *Contaminación atmosférica*. Obtenido de <https://www.agro.uba.ar/users/semmarti/Atmosfera/contatmosf.pdf>
- Sanitaria, G. (2019). *Efecto de las partículas de diámetro inferior a 2,5 micras (PM2,5) sobre los ingresos hospitalarios en niños menores de 10 años en Madrid*. Obtenido de https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112009000300005
- Servicio de Salud Pública. (2012). *Monóxido de Carbono*. Obtenido de https://www.atsdr.cdc.gov/es/phs/es_phs201.pdf