UCUENCA

Facultad de Ciencias Químicas

Carrera de Ingeniería Ambiental

Pronóstico de las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning

Trabajo de titulación previo a la obtención del título de Ingeniera/o Ambiental

Autores:

Dayana Mishel Ortiz Morocho

CI: 0605693183

Correo electrónico: daanaort96@gmail.com

Bryam Adrián Montesdeoca Jara

CI: 0105456362

Correo electrónico: bryammontesdeoca10-9@hotmail.com

Tutor:

Julio Danilo Mejía Coronel

CI: 0103638581

Cuenca - Ecuador

6 de enero - 2023

Resumen:

La contaminación del aire se ha convertido en uno de los principales problemas ambientales a nivel mundial debido a su afección tanto en el medio ambiente como en la salud en general. Los gobiernos tanto nacionales como internacionales han implementado esfuerzos para medir y controlar las emisiones de contaminantes al aire proveniente de fuente antrópicas instalando redes de monitorización atmosférica. Sin embargo, no todas las ciudades y países cuentan con estas herramientas de monitoreo. Por ello, el uso de las imágenes satelitales ha ido tomando fuerza en los últimos años ya que nos permite obtener información satelital de áreas que no cuentan con monitoreo terrestre y poder utilizar estos datos para fines de control, prevención e investigación. Por medio de dicha información podemos realizar análisis y modelado de las emisiones y comportamiento de los contaminantes atmosféricos.

Debido a la necesidad de poder prevenir a la sociedad y tomar medidas preventivas de las emisiones de contaminantes atmosféricos, la comunidad científica en los últimos años ha propuesto diferentes modelos matemáticos y modelos de aprendizaje no supervisado que permitan predecir las emisiones de los contaminantes atmosféricos. Para ello, es necesario tomar en cuenta las variables externas que afectan al comportamiento de los contaminantes dependiendo de la zona de estudio, ya que la ubicación geográfica, la topografía, y condiciones meteorológicas influyen directa o indirectamente en este comportamiento, por esta razón generalmente los investigadores diseñan modelos para regiones específicas. No existe un método para establecer qué variables meteorológicas deben ser usadas en la predicción de los contaminantes, los antecedentes a usar son los estudios previos realizados, observando los resultados obtenidos para saber las influencias de estas variables en el comportamiento de los contaminantes.

El presente trabajo propone dos modelos de predicción de la concentración de NO_2 y SO_2 para las tres ciudades más importantes del Ecuador Tomando como base la información de imágenes satelitales Sentinel-5P, Giovanni NASA y ERA 5. El primer modelo propuesto utiliza redes Neuronales Recurrentes utilizando el número de retrasos o variables ficticias creadas que se utilizan para encontrar relaciones entre la concentración y las variables meteorológicas, las cuales proporcionan información a la red neuronal para realizar la

predicción. Se propuso predecir la contaminación atmosférica hasta 5 días hacia adelante con el uso de diferentes estructuras buscando la mejor para el pronóstico. El segundo modelo propuesto utiliza el método de Random Forest teniendo en cuenta dos características importantes, la profundidad máxima de cada árbol y el número mínimo de muestras para considerarse Nodos Hoja. Estas dos características nos dan dos perspectivas acerca de los bosques aleatorios buscando el mejor modelo de predicción. Se puede decir que la predicción a través del algoritmo de Regresión de Random Forest fue el que mejor rendimiento R^2=0,98 mostró y las métricas de error MAPE, RMSE y PBIAS fueron más bajas en este método con valores de 7, 3,67, 0,68, respectivamente, haciendo énfasis en los distintos conjuntos de datos, la predicción para la ciudad de Cuenca fue la mejor seguida de la ciudad de Guayaquil que supera ligeramente a las predicciones de Quito. Esto demuestra que la predicción de la calidad del aire es efectiva mostrando resultados satisfactorios y abriendo puertas a nuevas investigaciones con la finalidad de poder prever las medidas de concentraciones de gases contaminantes al aire y así poder tomar decisiones preventivas tanto para la salud como el medio ambiente.

Palabras claves: Predicción. Redes neuronales recurrentes. Random forest. Contaminación. Ambiente.

Abstract:

Air pollution has become one of the main environmental problems worldwide due to its effects on both the environment and health in general. Both national and international governments have implemented efforts to measure and control air pollutant emissions from anthropogenic sources by installing atmospheric monitoring networks. However, not all cities and countries have these monitoring tools. For this reason, the use of satellite images has been gaining strength in recent years as it allows us to obtain satellite information from areas that do not have terrestrial monitoring and to be able to use this data for control, prevention and research purposes. Through this information we can perform analysis and modeling of emissions and behavior of atmospheric pollutants.

Due to the need to be able to prevent society and take preventive measures regarding the emissions of atmospheric pollutants, the scientific community in recent years has proposed different mathematical models and unsupervised learning models that allow predicting the emissions of atmospheric pollutants. For them it is necessary to take into account the external variables that affect the behavior of pollutants depending on the study area, since the geographical location, topography, and meteorological conditions directly or indirectly influence this behavior, for this reason researchers generally design models for specific regions. There is no method to establish which meteorological variables should be used in the prediction of pollutants, the background to be used are the previous studies carried out, observing the results obtained to know the influences of these variables on the behavior of pollutants.

The present work proposes two prediction models for the concentration of NO_2 and SO_2 for the three most important cities of Ecuador, based on information from Sentinel-5P, Giovanni NASA and ERA 5 satellite images. The first proposed model uses Recurrent Neural Networks using the number of lags or dummy variables created that are used to find relationships between concentration and meteorological variables, which provide information to the neural network to make the prediction. It was proposed to predict air pollution up to 5 days ahead with the use of different structures looking for the best one for the forecast. The second proposed model uses the Random Forest method taking into account two important characteristics, the maximum depth of each tree and the minimum number of samples to be considered Leaf Nodes. These two features give us two

perspectives about random forests looking for the best prediction model. It can be said that the prediction through the Random Forest Regression algorithm was the one that showed the best performance R2=0.98 and the error metrics MAPE, RMSE and PBIAS were lower in this method with values of 7, 3.67, 0.68, respectively. , emphasizing the different data sets, the prediction for the city of Cuenca was the best, followed by the city of Guayaquil, which slightly exceeds the predictions for Quito. This shows that the prediction of air quality is effective, showing satisfactory results and opening doors to new research in order to be able to anticipate the measurements of concentrations of polluting gases in the air and thus be able to make preventive decisions for both health and the environment.

Keywords: Prediction. Recurrent neural networks. Random forest. Pollution. Environment.

Índice

Contenido

Índice de Tablas

Índice de Gráficos

Índice de Ecuaciones

UCUENCA

Cláusula de Propiedad Intelectual

Dayana Mishel Ortiz Morocho, autora del trabajo de titulación "Pronóstico de las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autora.

Cuenca, 06 de enero del 2023

Dayana Mishel Ortiz Morocho

C.I: 0605693183

UCUENCA

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Bryam Adrián Montesdeoca Jara en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación "Pronóstico de las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 06 de enero del 2023

Bryam Adrián Montesdeoca Jara

C.I: 0105456362

UCUENCA

Cláusula de Propiedad Intelectual

Dayana Mishel Ortiz Morocho, autora del trabajo de titulación "Pronóstico de las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autora.

Cuenca, 06 de enero del 2023

Dayana Mishel Ortiz Morocho

C.I: 0605693183

UCUENCA

Cláusula de Propiedad Intelectual

Bryam Adrián Montesdeoca Jara, autor del trabajo de titulación "Pronóstico de las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning", certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, 06 de enero del 2023

Bryam Adrián Montesdeoca Jara

C.I: 0105456362

AGRADECIMIENTOS

En primer lugar deseo expresar mi agradecimiento a Dios y a la vida porque me ha permitido llegar hasta aquí desarrollando mis capacidades, dándome coraje y fortaleza para superar cada obstáculo.

A Mis padres José Ortiz y Susana Morocho, ustedes han sido siempre el motor que impulsa mis sueños y esperanzas, quienes estuvieron siempre a mi lado animándome a seguir, superando cada dificultad, sacrificándose para brindarme un futuro próspero y autosuficiente. Siempre han sido mis mejores guías de vida. Hoy cuando concluyo mis estudios, les dedico a ustedes este logro, amados padres, como una meta más conquistada. Orgullosa de haberlos elegido como mis padres y que estén a mi lado en este momento tan importante.

A la persona que con su apoyo y confianza me ha ayudado en cada paso de mi vida a la persona que considero una segunda madre, mi hermana Silvia Ortiz, quien con cada consejo y compañía ha estado en los momentos donde flaqueaba y celebrando mis pequeños triunfos buscando verme salir triunfante de cada batalla, no podría pedir mejor hermana en la vida.

A mi esposo Ronald Bravo, a quien amo tanto y agradezco por tenerme tanta paciencia, estar a mi lado en todo momento y por darme su amor todos los días, lo que me motiva a cumplir todo lo que me proponga. Me comprendiste y tuviste tolerancia conmigo para poder culminar esta etapa estudiantil para llegar a nuevos retos en nuestras vidas.

A mi alma mater, la Universidad de Cuenca y todos los docentes que de una u otra manera me han brindado un granito de arena para construir mi carrera profesional, Además, a mis compañeros con quienes compartimos tantos momentos y por las experiencias compartidas ya sea dentro o fuera de las aulas.

Finalmente me agradezco a mí, por a pesar de cada dificultad seguir adelante, con paso firme y mirando siempre al futuro con la fe para conseguir todos los sueños y realizar todos los proyectos planteados. Y con el apoyo de todas las personas que me aman he logrado llegar a este punto. ¡Gracias a todos!

Dayana

Agradezco en primera instancia al destino, la vida y a dios. Por darme la dedicación y la voluntad para culminar con todos los objetivos planteados en toda mi etapa universitaria y sobre todo por darme la fortaleza de luchar para conseguir el documento de titulación.

De igual manera a mis padres, por brindarme todas las herramientas posibles para sobrepasar los obstáculos presentes en toda mi vida. Y en especial por brindarme esa mano cuando las cosas estaban difíciles. Les agradezco por esos consejos, por esas enseñanzas que marcaron en mi para ser lo que hoy soy como persona. A mi madre Marlene Jara que en mis peores momentos ha sabido sacar de mi esa chispa que todos tenemos para brillar y para seguir adelante a pesar de tener muchos obstáculos en frente. A mi padre Gonzalo Montesdeoca que, aunque la distancia nos separa siempre ha estado ahí para mí de una u otra manera guiándome por el mejor camino y enseñándome que con trabajo y dedicación todo es posible. Sin ustedes todo esto no hubiera sido posible.

Además, todo esto fue posible con la ayuda de muchas personas magníficas con las que coincidí y compartí durante mucho tiempo en la universidad. Entre altos y bajos me han demostrado el significado del término amistad. Agradezco que siempre han estado conmigo ya sea en lo más insignificante o en este caso lo más grande.

Bryam

1. INTRODUCCIÓN

La contaminación del aire se remonta a cuando el hombre descubrió el fuego, sin embargo, la contaminación provocada por éste fue mucho menor a la causada por fuentes naturales. A través de los años, el hombre se desarrolló en comunidades permanentes dando paso al desarrollo agrario y posteriormente a procesos industriales altamente avanzados provocando efectos adversos en el ambiente y la salud (Michél et al., 2013). Ante esta problemática el año 1972 se celebró en Estocolmo la Primera Conferencia sobre el Ambiente Humano de la Organización de las Naciones Unidas donde se persuadió a muchos países a desarrollar la legislación necesaria con la finalidad de limitar las emisiones de contaminantes químicos tóxicos al ambiente, del mismo modo que se implementen nuevas tecnologías y políticas con la misma finalidad, que dio como resultado la reducción de problemas ambientales de contaminación industrial (Romero et al., 2006).

En las últimas décadas, la contaminación atmosférica ha sido uno de los temas de investigación más importantes en distintos estudios ambientales que tiene un alcance a nivel regional y Global (Ramírez 2017), esto se debe en gran medida al rápido y desordenado crecimiento urbano, a la industrialización y a la combustión (Lacasaña-Navarro et al. 1999). Según World Health Organization WHO (2011) alrededor del 99 % de la población mundial respira aire que supera los límites permisibles, destacando que los países de bajos y medianos ingresos sufren las exposiciones más elevadas causando afecciones a la salud como enfermedades respiratorias y cardiacas.

Las consecuencias por la contaminación ambiental son muchas, entre las más importantes se conoce como efecto invernadero el cual es provocado por gases en la atmósfera que provoca un calentamiento adicional de la temperatura de la tierra y la destrucción de la capa de ozono, la cual tiene una función muy importante que es la absorción de la radiación ultravioleta procedente del espacio exterior lo que permite la vida en la tierra (Martínez y Díaz de Mera, 2004).

Los gases emitidos al aire son llevados a la tropósfera que es la capa más próxima a la tierra en donde se da lugar a la mayor parte de procesos de oxidación de las sustancias presentes. Esta oxidación se produce por la presencia de óxido de nitrógeno y radiación solar que da como resultado el ozono. Entre las partículas emitidas al aire que causan

preocupación son el monóxido de carbono (CO), dióxido de nitrógeno (NO_2) y dióxido de azufre (SO_2) (Martínez, 2021).

Las grandes ciudades en particular son grandes emisoras antropogénicas, sin embargo, a pesar de los estudios que cuantifican y comprenden estas emisiones en zonas urbanas carecen de un análisis de alta resolución (Kort et al., 2012). Existen tres métodos para la estimación y análisis de los contaminantes atmosféricos como son los monitoreos terrestres, estimación de datos mediante satélites y modelos atmosféricos (Marlier et al., 2016).

Con el desarrollo tecnológico, las nuevas plataformas digitales y el lanzamiento de nuevos satélites de teledetección de alta resolución, las observaciones de emisiones de contaminantes atmosféricos se han utilizado cada vez más para el monitoreo de la atmósfera debido a la ventaja que brindan un muestreo espacial continuo con cobertura global, incluso observando áreas donde las estaciones de monitoreo terrestres tiene poca o ninguna cobertura que ayude a monitorear las emisiones con suficiente precisión (Park et al. 2021).

Tal es el caso que en 13 de octubre del 2017 fue puesto en marcha la misión Copernicus Sentinel-5 Precursor dedicada a monitorear la atmósfera, esta misión consiste en un satélite que lleva el instrumento Tropospheric Monitoring Instrument (TROPOMI) el cual es un espectrómetro de rejilla a bordo del satélite Sentinel-5 Precursor (S5P) que mide las trazas de gases atmosféricos diarios. Entre ellos Monóxido de carbono (CO), Dióxido de Azufre (SO_2), Dióxido de Nitrógeno (NO_2), Ozono (O_3) y Formaldehído (HCHO) (Agencia Espacial Europea [ESA], 2022).

De la misma manera, el análisis de imágenes satelitales nos permite obtener datos meteorológicos mensuales y horarios de un gran número de variables atmosféricas oceánicas y terrestres (ECMWF, 2022). El centro de Servicios de Información y Datos de la Ciencia de la Tierra Goddard de la NASA a través de un sistema de gestión de datos World Wide Web Geovanni ha proporcionado una interfaz intuitiva y receptiva para visualizar datos de sensores múltiples para descargar imágenes y datos en múltiples formatos. Además, permite el análisis de datos en la investigación de eventos y procesos geofísicos con datos de teledetección (Berrick et al., 2009).

Existen muchas maneras de aplicar datos de concentración de contaminantes a la atmósfera con la finalidad de estimar o predecir concentraciones futuras, para ello, Aguirre et al., (2006) explica que los modelos basados en el uso de técnicas estadísticas para establecer relaciones funcionales entre las variables de entrada y salida son oportunos para el objetivo programado. El uso de la analítica de aprendizaje y el uso de Redes Neuronales Recurrentes son modelos matemáticos-computacionales de tratamiento de información cuyo origen es la simulación del cerebro humano.

En el presente trabajo se pronosticó la concentración de dióxido de nitrógeno (NO_2) y dióxido de azufre (SO_2) en Ecuador, a partir de datos de concentración de gases obtenidos de imágenes satelitales Sentinel-5P en un periodo del año 2019 y 2020 como base de datos. Además, se tomó datos meteorológicos diarios como relación con el comportamiento de estos gases, los mismos que fueron obtenidos de las plataformas de la NASA Y ERA 5. Los datos obtenidos fueron tratados y correlacionados mediante Redes Neuronales y Random Forest con la finalidad de pronosticar la concentración de los gases hasta 5 días seguidos después de haber sido determinada la contaminación por estos gases a la atmósfera, con la finalidad de tomar acciones preventivas enfocadas hacia la ciudadanía y que las mismas puedan precautelar su salud y movilidad diaria.

2. PLANTEAMIENTO Y FORMULACIÓN DEL PROBLEMA

El impacto de la contaminación del aire es un tema muy esencial para el clima y el medio ambiente. La atmósfera es fundamental para la vida, por lo que sus alteraciones representan un gran impacto en los seres humanos, otros organismos y el planeta en su conjunto. Pero más allá de eso, los cambios que se producen en la composición química de la atmósfera pueden alterar el clima, generar lluvia ácida o destruir la capa de ozono, todos fenómenos de importancia global.

Según Romero et al., (2010) dentro de las fuentes actuales más relevantes de contaminación atmosférica se encuentran los vehículos, las industrias y el consumo de leña. Además, en las zonas costeras existe una elevada concentración de contaminantes atmosféricos provenientes de los buques de carga (Zambrano, 2014), como también la quema de basura por parte de los botaderos y la explotación de canteras en la zona (Espín, 2011). El desarrollo industrial ha alcanzado en los últimos años un gran avance, lo que ha contribuido al aumento de las emisiones atmosféricas y estas son emitidas continuamente a través de sus chimeneas Moscoso et al., (2018). Peña (2018) menciona que a nivel nacional se ha registrado altas concentraciones de contaminantes atmosféricos, estos datos informan sobre el comportamiento de micro partículas y gases como Ozono (O_3), Dióxido de Azufre (SO_2), Dióxido de Nitrógeno (NO_2), Monóxido de Carbono (CO), Material Particulado menor a 10 micras (PM_{10}) y Material Particulado menor a 2,5 micras ($PM_{2,5}$).

Según Coronel y Marzo (2017) la contaminación del aire en América Latina es un asunto preocupante en las ciudades emergentes de la Región, como consecuencia de la quema de combustibles fósiles, entre las fuentes es el deficiente planeamiento de transporte, la generación de energía, el sector industrial y manufacturero y utilización de combustibles de mala calidad. Como consecuencia de la emisión de estos contaminantes, se produce el deterioro de la atmósfera y el medio ambiente. Llanque, (2003) menciona que entre los efectos que se producen son la radiación solar directa, olas de calor, la formación de lluvia ácida, la formación de nubes nocivas que cubren las ciudades y el fenómeno de inversión térmica.

Por otro lado, los efectos generados a la salud por la concentración de contaminantes atmosféricos ha sido un tema de gran estudio para los científicos de la actualidad, esto

debido a las consecuencias que generan a corto y largo plazo ocasionando en la mayoría de los casos enfermedades respiratorias y cardiovasculares (Brunekreef y Holgate, 2002).

Estudios publicados en los últimos años han relacionado el aumento de asma en niños, ausencia en escuelas y hospitalización de niños que viven cerca de vías con alta circulación vehicular (ANMM, 2015). Según Montero et al., (2020) existe una elevada incidencia de rinitis alérgica en aproximadamente el 20 % de recién nacidos en la ciudad de Riobamba. En donde el 58 % de los casos se dan en niños los cuales habitan zonas urbanas, tomando en cuenta que las investigaciones realizadas en Chimborazo a través del plan nacional de la calidad del aire registran 31 764 casos de infecciones respiratorias agudas. Y en la ciudad de Cuenca se registró que 101 personas fallecieron en el año 2012 por enfermedades cardiopulmonares y cáncer de pulmón, ocasionadas por niveles elevados de contaminantes atmosféricos (Palacios y Espinoza, 2014).

Actualmente, Ecuador es un país el cual se encuentra en vías de desarrollo industrial, por lo que es necesario estimar la concentración de emisiones contaminantes a largo plazo, de esta manera evitar ya sea de forma directa o indirecta afecciones generadas por las elevadas concentraciones de contaminantes atmosféricas.

Existen varios procesos para predecir las concentraciones de contaminantes en la atmósfera. Sin embargo, uno de los métodos más usados por su simplicidad para transformar grandes series de datos y por su eficiencia en los resultados son las técnicas de lenguaje no supervisado Machine Learning. Por esta razón en este trabajo se plantea un modelo para pronosticar la concentración de los contaminantes NO_2 y SO_2, obtenidos de imágenes satelitales del tiempo correspondiente, comparados con los valores reales obtenidos de imágenes satelitales posteriores para las tres principales ciudad del Ecuador Quito, Guayaquil y Cuenca con la finalidad de alertar a la población de posibles riesgos en la salud y advertir a las empresas que contribuyen a las emisiones para que tomen las medidas pertinentes.

3. OBJETIVOS

3.1. Objetivo General

- Pronosticar las concentraciones de SO_2 y NO_2 en Ecuador a partir de imágenes satelitales Sentinel 5P, mediante técnicas de Machine Learning.

3.2. Objetivos Específicos

- Obtener datos de concentración de contaminantes atmosféricos a partir de imágenes satelitales.
- Procesar las imágenes satelitales obtenidas para establecer una línea base de investigación.
- Construir modelos para el pronóstico de emisiones de concentraciones de contaminantes atmosféricos mediante Redes Neuronales Recurrentes y Bosques Aleatorios (Random Forest).
- Validar los datos obtenidos de la predicción para de esta manera determinar la eficiencia de los métodos mediante las métricas de rendimiento y error.

4. MARCO TEÓRICO

4.1. Antecedentes y Estudios previos

En la investigación del caso de interés e intentando encontrar una definición del problema que sea ajustada a la verdad enfocada y plausible de desarrollar se ha revisado la literatura que existe por medio de base de datos, intentando encontrar precedentes de estudios, investigaciones o artículos semejantes o involucrados que nos posibilite formar una base sólida para el desarrollo de la presente investigación de forma que las conclusiones y resultados logrados sean pertinentes, permitiendo la generación de conocimientos y valor añadido para los interesados.

Investigaciones alrededor del mundo han demostrado el creciente interés en la aplicación de lenguaje no supervisado para estimaciones de emisión de contaminantes a la atmósfera. Sanjuán de Caso (2020) en un estudio donde intentó predecir las concentraciones de contaminantes mediante técnicas de Machine Learning en la ciudad de Madrid utilizó Redes Neuronales Long Short Term Memory (LSTM) y Gated Recurrent Unit (RGU). La base de datos de concentraciones de contaminantes se obtuvo de las estaciones de control de calidad del aire y variables meteorológicas como variables independientes. Utilizó Python como el lenguaje de programación buscando el método más preciso. Como resultado se obtuvo que las Redes Neuronales LSTM donde utilizó entre 5 a 10 neuronas fue más precisa, además se aplicó la estacionalidad demostrando que en las estaciones de primavera y verano muestra predicciones más precisas caso contrario sucede en otoño e invierno donde los valores decaen.

Además, estudios recientes realizados en América Latina muestra el abordaje reciente en la que Pedraza (2019), desarrolló un prototipo para demostrar el rendimiento y eficacia que posee el modelo de redes neuronales utilizando machine Learning aplicado a la predicción del material particulado ($PM_{2.5}$) en la ciudad de Bogotá. Los datos fueron obtenidos de la estación de monitoreo de la calidad de aire de la localidad de Kennedy. En la modelación se utilizaron varias librerías de Python donde se dio como resultado positivo con errores del 3,56 % y 3,67 % demostrando que el método basado en redes neuronales para estimar

las concentraciones de $PM_{2,5}$ son un método de aplicación para casos futuros de predicción.

El trabajo realizado por Jacinto (2019) demuestra la factibilidad del uso de técnicas de inteligencia artificial utilizando redes neuronales del tipo multicapa, en donde se logró predecir la presencia de contaminantes atmosféricos $PM_{2.5}$ en Carabayllo-Lima. Los datos se obtuvieron de la estación meteorológica de Carabayllo, en la metodología se aplicó sobre tres algoritmos de retropropagación y se utilizó una capa oculta para analizar dos modelos artificiales obteniendo un error porcentual medio de -0,1089 % lo cual apoya la teoría de aplicación para predecir contaminantes atmosféricos mediante esta metodología.

Por otra parte, Herrera (2019) desarrolló un sistema de predicción del índice de calidad del aire de dióxido de carbono en la ciudad de Santo Domingo mediante Redes Neuronales Artificiales, donde para la obtención de los datos utilizó sensores de medición de gases en el aire y un controlador para su almacenamiento, a través del software Matlab modeló la red neuronal de tres capas con una capa de entrada de 4 neuronas, una capa oculta de 10 neuronas y la capa de salida con 1 neurona mediante la codificación de algoritmos. Los valores obtenidos de la simulación fueron comparados con los valores reales demostrando un error promedio de 1,47 de los casos validados, mostrando que es un modelo enormemente aceptado. Además, se predijeron todos los meses del 2019 a 2022 para verificar la efectividad de la predicción a corto y largo plazo donde se señaló un aumento importante de 0,8 ppm para el año 2022 tomando en cuenta las ocupaciones humanas que se desarrollan en esta región.

Desde otra perspectiva, Gavilánez (2021) demostró la efectividad de las Redes Neuronales Artificiales (RNA) considerando el algoritmo de retropropagación en cascada utilizadas para la estimación de las concentraciones elevada de dióxido de carbono en interiores de laboratorios químicos de la Universidad técnica de Ambato. Para la modelación utilizó el algoritmo Back Propagation que se basa en una regla de corrección instantánea relacionando los datos medidos con diversos factores que afectan la calidad del aire. Como resultados se observó que el modelo posee un rendimiento de aproximadamente el 80 % y en donde la validación demostró que se puede obtener errores de hasta un 0,99557. Y de esta manera se demuestra la adaptabilidad que posee el modelo para ser aplicado con otros contaminantes del aire en donde se pueda tener ambientes de buena presión.

Estudiando el documento de (Liu et al., 2021) el cual realiza una red neuronal recurrente utilizando el modelo de unidad recurrente cerrada para pronosticar la concentración de PM 2,5 en un periodo de 24 horas. El estudio se realizó en Beijing, en el cual para el análisis se tomó en cuenta 35 estaciones de la calidad del aire. El método utiliza mapas autoorganizados geográficos, en donde organiza en grupos las estaciones de monitoreo de varias zonas agrupándolas por coordenadas geográficas. Los resultados obtenidos validan el modelo. Se analizó el coeficiente de determinación (R^2), el error relativo medio (MRE) y el error absoluto medio (MAE), dando como resultados en la estación de Gucheng valores de 0,35 para el coeficiente de determinación, 0,79 para el MRE y 16,1 para el MAE.

En el documento realizado por (Li et al., 2018) se realiza un método en línea basado en Random Forest para el pronóstico de la concentración de NO_2, SO_2 Y $PM_{2.5}$ con 24 horas de antelación. En el estudio se utilizan tres métricas de evaluación. Se utilizó el error absoluto medio (MAE), error cuadrático medio (RMSE) y accuracy (Acc), en donde para determinar la eficiencia del modelo se compara con otros modelos como un modelo predictivo híbrido, una red neuronal profunda pre entrenada por una red de creencias profundas. En donde los resultados del estudio determinaron que el método propuesto supera los resultados ya que este es capaz de predecir la concentración de contaminantes con mayor precisión.

Por otra parte (Vu et al., 2019) desarrolló un modelo de aprendizaje automático basado en Random Forest con el cual se estima las concentraciones de $PM_{2.5}$ en la ciudad de Lima, Perú. Para el desarrollo del modelo se obtuvieron los datos de 16 estaciones de monitoreo desde el año 2010 hasta el 2016. Además de esto se combinaron datos satelitales como la profundidad óptica de aerosoles, datos de campos meteorológicos del centro Europeo para pronósticos meteorológicos a medio plazo, parámetros del modelo weather research and forecasting y variables de uso de la tierra para ajustar el modelo. Se estimó la media de concentración para el $PM_{2.5}$ medida en tierra la cual alcanzó un valor de 24,7 $\mu g/m^3$ mientras que para la validación cruzada el valor de $PM_{2.5}$ fue de 24,9 $\mu g/m^3$, estableciendo una diferencia de 0,09 $\mu g/m^3$ (desviación estándar de 5,97 $\mu g/m^3$) por lo tanto el modelo representa una enorme concordancia entre valores por lo que el modelo puede ser tomado como positivo debido a la veracidad de los resultados.

Con respecto a Random Forest no se ha podido encontrar estudios varios en el país ni cercanos a nuestra área de estudio, un estudio realizado en Madrid por (Pérez, 2021)

busca obtener un modelo predictivo para los contaminantes emitidos en dicha ciudad y la influencia de las variables meteorológicas y la reducción del tráfico como relación con el contaminante en donde para generar el Bosque Aleatorio tomo en cuenta las características de construcción entre ellas, la profundidad máxima de cada árbol y el número mínimo de cada nodo hoja donde concluyó que Random Forest capta mejor la variabilidad de las concentraciones observadas. Además, concluyó que el tráfico y los datos meteorológicos se pueden tratar como variables independientes.

Singh et al., (2021) realizaron un estudio similar en la región Urbana de Nueva Delhi en la India donde utilizo el algoritmo de Bosque aleatorio para predecir la concentración de varios contaminantes considerando la misma estructura que el estudio anterior donde se obtuvieron resultados satisfactorios ya que utiliza muchos algoritmos de árbol para encontrar el resultado, por lo que obtenemos una solución más precisa con este algoritmo.

En Ecuador los estudios en Lenguaje no supervisado se han ido incrementando recientemente. Sin embargo, la utilización de imágenes satelitales relacionadas con técnicas de Machine Learning han sido escasas.

4.2. Base Teórica

4.2.1. La atmósfera

Según Venegas y Mazzeo (2012) la atmósfera es una capa gaseosa de aproximadamente 10000 km de espesor que rodea la tierra, donde se producen todos los fenómenos climáticos y meteorológicos que afectan al planeta, además regula la entrada y salida de energía de la tierra. Es la principal defensa que tienen las diferentes formas de vida de la incidencia de las radiaciones provenientes del espacio exterior, especialmente del sol.
La atmósfera muestra variaciones verticales de temperatura producida por diferentes procesos radiactivos, dinámicos y químicos en las diferentes alturas de la atmósfera, por esta razón se divide en cuatro capas separadas por tres zonas de transición, las cuales se denominan tropósfera, estratósfera, mesosfera y termósfera y las zonas de transición tropopausa, estratopausa y mesopausa respectivamente (Sáenz, 2016).

La Tropósfera es la capa que se encuentra en contacto con la tierra por ende es la más cercana a nuestro planeta siendo la capa más densa de la atmósfera y en ella se originan los fenómenos atmosféricos. La dinámica de esta capa hace posible el funcionamiento de los ecosistemas terrestres y acuáticos, factores como la precipitación o la radiación solar que llega al suelo puede verse afectada por el contenido de aerosoles en la troposfera. Concentra la mayor parte de oxígeno y vapor de agua, además actúa como regulador térmico del planeta haciendo posible la vida en el (Morales et al., 2001).

4.2.2. Contaminación Atmosférica

Se puede definir como contaminación del aire a la introducción de partículas y sustancias químicas o biológicas en la atmósfera que causan daño o afectación a los seres vivos. Los contaminantes dañan nuestro medio ambiente ya sea aumentando los niveles por encima de lo normal o al introducir sustancias tóxicas dañinas (Hutton, 2011). Esto afecta principalmente a las personas que viven en grandes ciudades, siendo las emisiones de las carreteras el mayor contribuyente al deterioro de la calidad del aire (Manisalidis et al., 2020).

Además, la contaminación del aire puede influir en la calidad del suelo y de las masas de agua al contaminar las precipitaciones que caen en el agua y el suelo modificando la química del suelo y causando desequilibrios en el pH del mismo (Manisalidis et al., 2020). Además, puede afectar a la salud al estar expuesto a altas concentraciones de contaminantes durante prolongados periodos de tiempo. En el gráfico 1 podemos apreciar el ciclo de los contaminantes que empiezan por la emisión desde una fuentes natural o antropogénica y finaliza nuevamente en la tierra o agua en forma de otros contaminantes alterando el equilibrio químico y natural del ambiente.

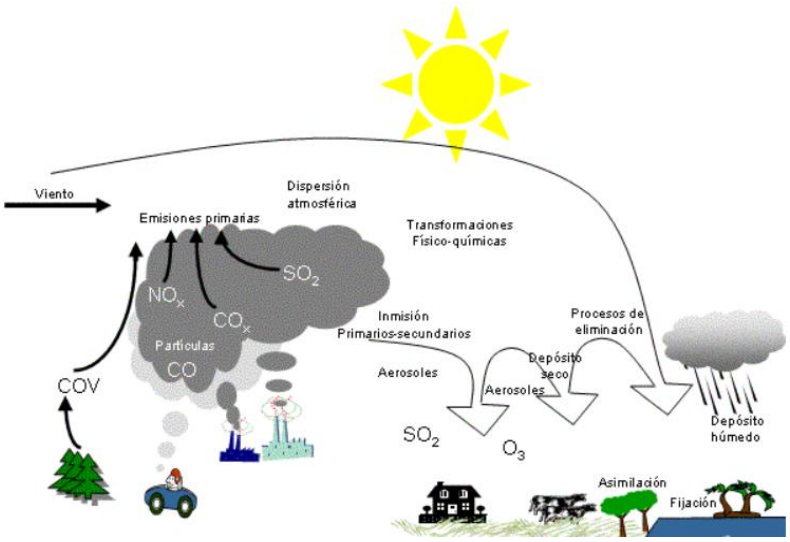

Gráfico 1. Ciclo de la contaminación atmosférica. (Llosa, 2010)

4.2.2.1. Fuentes de contaminación atmosférica

Las fuentes de contaminación en la atmósfera varían según su procedencia, esto debido a que en un principio las emisiones generadas fueron el factor esencial para que la vida en la tierra tenga procedencia. Existen fuentes naturales de contaminación al aire como son erupciones volcánicas, descomposición biológica, incendios, marismas y materiales con altos niveles de radiación. Y existen fuentes antrópicas como las emisiones generadas por vehículos, quema de combustibles fósiles, actividad agrícola, plantas de energía térmica o industrias (Puri et al., 2017).

Actualmente, en todo el mundo la contaminación del aire representa una nueva problemática debido a los impactos toxicológicos en el ambiente y en general en la salud humana. Según la Organización Mundial de la Salud los principales contaminantes presentes en el aire son las partículas en suspensión, el ozono almacenado en la tropósfera, el monóxido de carbono, el plomo, los óxidos de azufre y óxidos de nitrógeno. En donde se ha demostrado que su procedencia corresponde en mayoría a los vehículos de motor y a los procesos industriales (Ghorani-Azam et al., 2016). En China se estimó que los niveles elevados de contaminación de óxidos de nitrógeno son generados por el transporte automotriz (15 a 25 %), centrales eléctricas las cuales queman combustibles fósiles (30 a 50 %) y centros industriales (25 a 35 %) (Rohde y Muller, 2015).

Tal es el caso que en una gran parte de EE. UU. la contaminación del aire es un tema complejo debido a las regularizaciones en donde como fuentes principales tenemos a las emisiones del tráfico y algunas industrias las cuales generan energía. Mientras que hablando de partículas en suspensión las cuales son fuentes principales de impacto a la salud humana, tenemos que en el este de EE. UU., Europa, Rusia y el este de Asia la producción agrícola es un factor representativo para la elevada concentración de contaminantes (Lelieveld et al., 2015).

En otras regiones, como en los países asiáticos los cuales poseen una densidad poblacional elevada, continúan afectados por la mala calidad del aire, y se pronostica que en años venideros estos problemas aumenten. La combinación entre poblaciones muy grandes y la alta contaminación representan un elevado costo para la salud pública (Jerrett, 2015).

4.2.2.2. Clasificación de contaminantes atmosféricos.

Los contaminantes se pueden clasificar en contaminantes primarios y secundarios. Los contaminantes primarios son emitidos directamente desde la fuente, entre ellos, el Dióxido de Azufre (SO_2), Óxido de Nitrógeno (NO_x), Monóxido de Carbono (CO), Plomo (Pb), compuestos orgánicos y partículas en suspensión (Martínez y Díaz de Mera, 2004) o como emisión de procesos antropogénicos, como el monóxido de carbono que es emitido del escape de los vehículos o el dióxido de azufre liberado de procesos industriales (Hutton, 2011).

Los contaminantes secundarios se forman cuando ocurre una reacción química en la atmósfera que involucra a un contaminante primario en el aire, como el dióxido de azufre que puede oxidarse para formar ácido sulfúrico que luego puede reaccionar con amoniaco y formar sulfato de amonio (Farrow et al., 2020), los contaminantes secundarios incluyen: el ácido nítrico (HNO_3), ácido hiposulfuroso (H_2SO_2) y el ozono (O_3) (Páez, 2022).

Para este estudio haremos énfasis en el Dióxido de Nitrógeno y Dióxido de Azufre puesto que son los contaminantes a considerar para los objetivos propuestos.

4.2.2.2.1. Dióxido de Nitrógeno

El NO_2 es un gas contaminante principalmente urbano que está altamente correlacionado con la población y el tráfico de vehículos ya que es emitido por los motores de los mismos (Anenberg et al., 2020). Es un contaminante ampliamente prevalente de fuentes naturales y antropogénicas que puede causar irritación respiratoria cuando es inhalado en altas concentraciones (Hesterberg et al., 2009). Según estudios realizados se ha demostrado que la exposición a largo plazo a altos niveles de dióxido de nitrógeno causa enfermedades pulmonares crónicas, deterioro del sentido del olfato y síntomas documentados como irritación de ojos, garganta y nariz. De la misma manera afecta a la vegetación debido a que influye en la eficiencia de crecimiento de las plantas y su rendimiento (Manisalidis et al., 2020).

4.2.2.2.2. Dióxido de Azufre

El dióxido de azufre (SO_2) es un gas incoloro con un fuerte olor que se produce durante la combustión de combustibles fósiles, la fundición, producción de ácido sulfúrico, incineración de desechos y la producción de azufre elemental (Katulski et al., 2011). Los niveles peligrosos de contaminación se encuentran comúnmente cerca de las centrales eléctricas de carbón, en las refinerías y en áreas designadas por la industria pesada que se encuentra principalmente en Rusia, India y China en donde se registran concentraciones muy elevadas (Dahiya et al., 2020).

Este contaminante es un irritante sensorial por lo que causa irritación respiratoria bronquitis, producción de moco y broncoespasmo (Manisalidis et al., 2020). Además, puede provocar accidentes cerebrovasculares, enfermedades cardíacas, asma, cáncer de pulmón, muerte prematura (Dahiya et al., 2020) y problemas de fertilidad (Carré et al., 2017).

4.2.2.3. Efectos de la contaminación atmosférica en la salud.

La Comisión Europea estimó que por causa de la contaminación del aire mueren prematuramente 310.000 personas en 11 países europeos. Además de que se estimó que

por partículas en suspensión en el aire existen alrededor de 40.000 personas las cuales fallecen de forma prematura, como consecuencia la esperanza de vida media se reduce en al menos nueve meses (Robinson, 2005). Los efectos negativos generados a la salud humana producto de la contaminación atmosférica incluyen estrés oxidativo, autofagia, interrupción de la barrera epitelial respiratoria y las vías de señalización celular, inmunidad celular desregulada, mutaciones epigenéticas y la infiltración de células inflamatorias (Guan et al., 2016).

En Irán se determinó que, entre las 10 enfermedades causantes de muerte prematura por mala calidad del aire, cuatro están relacionadas a la arteriosclerosis. Además de que ciertos contaminantes afectan las vías respiratorias e irritan los pulmones. Los grupos más vulnerables son personas sensibles como los niños, mujeres en etapas de gestación, ancianos y personas con problemas de salud. Los efectos a corto plazo van desde molestias respiratorias o tos hasta alergias y estornudos. Y a largo plazo en presencia de ciertos contaminantes los efectos pueden ser más perjudiciales ocasionando cáncer en los pulmones (Hosseini y Shahbazi, 2016).

Se puede señalar que las consecuencias de esta contaminación se ven más afectadas en países con poco o nulo desarrollo económico, siendo estos países poseedores de algunos de los niveles más altos de contaminación a nivel mundial, ya que el tratamiento y la aplicación de las regulaciones son escasas en la mayoría de los casos. Una mala calidad de aire se simplifica en mayores costos para cuidado de la salud relacionadas con enfermedades cardiovasculares, respiratorias por lo que esto afecta en la productividad y ocasiona una menor esperanza de vida. A pesar de esto, los valores cuantificados de los costos necesarios para este fin aún no se comprenden bien debido a la falta de datos de gastos médicos y falta de análisis de relaciones entre la dosis y la respuesta al contaminante (Xia et al., 2022).

4.2.2.4. Variables que intervienen en la contaminación atmosférica.

Generalmente los problemas de contaminación en una zona específica no suelen ser ocasionada por los niveles elevados de descarga de contaminantes, sino que estos suelen

ser provocados por factores meteorológicos desfavorables. Además, el paisaje urbano es un factor influyente para esto ya que es donde se desarrollan todas las actividades que generan contaminación. Estas condiciones desfavorables son producto de una estrecha relación con la poca capacidad de la atmósfera de la zona para dispersar y transportar los contaminantes desde la fuente de emisión a varias zonas aledañas (Han et al., 2015).

Existe un gran interés entre la variación espacial y temporal de la concentración de gases contaminantes por lo que (Li et al., 2014), realizó un estudio en el cual se compara el índice de contaminación del aire con los factores meteorológicos en Guangzhou, China entre el año 2001 y el 2011. Concluyendo que las condiciones meteorológicas de la zona son importantes fuerzas impulsoras para determinar la concentración de gases. Pero sobre todo se descubrió que la presión en la zona es un factor determinante, esto debido a que cuando el lugar está dominado por un sistema de bajas presiones el aire asciende provocando un arrastre de los contaminantes a mayores altitudes facilitando de esta manera la dilución de los mismos. Por lo contrario, si la zona está dominada por un sistema de altas presiones, el aire se estabiliza de tal forma que es muy difícil la dispersión de contaminantes en el aire.

Así mismo, el análisis procedente de (Yang et al., 2019), determinó que uno de los factores más influyentes son las condiciones meteorológicas siendo así que las condiciones climáticas en verano representan características más propicias para la eliminación y la dispersión de los contaminantes en las zonas analizadas, en donde el viento, la precipitación y la radiación fueron los factores más influyentes.

4.2.3. Monitoreo de la calidad del aire

Actualmente la población en su mayoría vive en áreas urbanas las cuales en los últimos años son más vulnerables a la contaminación del aire y los problemas que se derivan de ello, debido a que esta tiene impacto negativo en las economías locales y nacionales (Loenen et al., 2021). El control de la calidad del aire mediante diferentes métodos, calibraciones y mediciones de los sitios y ambientes tanto exteriores como interiores son de importancia crucial ya que se utiliza para crear modelos para la evaluación de la

exposición, el mismo que sirve para la investigación de la contaminación del aire en la salud ambiental (Phuleria, 2013).

4.2.4. Monitoreo de la calidad del aire con teledetección satelital

En la actualidad el uso de la teledetección ha avanzado a grandes pasos desde la interpretación de fotografías hasta el análisis de imágenes satelitales debido a que los sensores pueden proporcionar datos de la energía emitida, reflejada y/o transmitida desde todas partes del espectro electromagnético (Anyamba et al., 2015).

El monitoreo atmosférico basado en satélites es clave para comprender las condiciones del aire y sus tendencias a escala regional y global, proporcionando información cuantitativa sobre las emisiones y el transporte de contaminantes (Páez, 2022). Uno de los sensores utilizados en la actualidad para poder determinar las emisiones de concentraciones de gases es el Instrumento de Monitoreo Troposférico (TROPOMI) del Satélite Sentinel-5 Precursor (Veefkind et al., 2012).

4.2.5. Satélite Sentinel-5P

Sentinel-5 Precursor de la ESA (Agencia Espacial Europea) es la primera misión de monitoreo atmosférico de Copernicus que tiene la finalidad de reducir la brecha de datos entre el satélite Envisat y el lanzamiento de Sentinel-5. El satélite lleva un instrumento TROPOMI más avanzado para mediciones de bandas espectrales ultravioleta-visible (270–500 nm), infrarrojo cercano (675–775 nm) e infrarrojo de onda corta (2305–2385 nm). Esto significa que puede detectar diferentes contaminantes como NO_2 , O_3 , CH_2O, SO_2 , CH_4, CO y aerosoles con una precisión sin precedentes (Zheng et al., 2019).

Existen diferentes productos de datos asociados con los tres niveles de procesamiento de TROPOMI. El nivel 0 es la telemetría satelital sin procesar ordenada por tiempo sin superposición temporal, estos datos no están disponibles para el público. El nivel 1B son datos geolocalizados en la atmósfera y corregida radiométricamente Radiancias de Tierra como irradiancias solares. Los datos de nivel 2 (L2) incluye columnas totales

geolocalizadas de ozono dióxido de azufre, óxidos de nitrógeno, monóxido de carbono, formaldehído y metanol, columnas de ozono troposférico, perfiles verticales geolocalizados de ozono, e información geolocalizada de nubes y aerosoles los cuales están agrupados por tiempo no por longitud y latitud. Finalmente, los datos de Nivel 3 (L3) son aquellos que manteniendo una sola cuadrícula por órbita convierte los datos L2 en datos L3 (*Sentinel-5P Mission - Sentinel Online*). Además, existen dos tipos de servicios, los datos en tiempo casi real (NRTI) que están disponibles hasta 3 horas después de posteriores a la detección y los datos fuera de línea (OFFL) que pueden estar disponibles hasta 12 horas después y para las columnas de metano, ozono troposférico y dióxido de nitrógeno corregido tardaría hasta 5 días después de ser detectados (van Geffen et al., 2021).

4.2.6. Google Earth Engine

Google Earth Engine (GEE) es una plataforma de computación en la nube diseñada para almacenar y procesar grandes cantidades de datos geoespaciales para un posterior análisis (Mutanga y Kumar, 2019). A partir de la cual se pueden producir datos sistemáticos o implementar aplicaciones interactivas respaldadas por los recursos de GEE (Gorelick et al., 2017). El archivo actual incluye los datos de varios satélites como Landsat, Sentinel, MODIS entre otros (Jin, 2020) así como conjuntos de datos basados en sistemas de información geográfica demográficos meteorológicos, modelos digitales de elevación y capas de datos climáticos (Mutanga y Kumar, 2019).

GEE está compuesto por cuatro elementos principales, el primero es la infraestructura de Google permitiendo hacer análisis en paralelo con cerca de 10000 CPUs, el segundo es el dataset o acervo de datos, mismo que se actualiza a medida que se toman nuevas imágenes creando de esta manera un gran catálogo de datos geoespaciales, el tercer elementos es la API (Application Program Interface) las misma que consiste en una serie de comandos o funciones en lenguaje JAVA permitiendo una programación sencilla al desarrollar algoritmos para investigación, por último, el cuarto elemento es el Code Editor donde a través de los códigos o scripts permite buscar la información en la nube y visualizarlos o procesarlos para distintos fines (Perilla y François, 2020).

4.2.7. Datos Sentinel-5P disponibles en Google Earth Engine.

GEE utiliza un modelo de datos simple basado en bandas ráster cuadriculadas en 2D. Los píxeles de una banda individual deben ser consistentes en términos del tipo de datos, resolución y proyección. No obstante, las imágenes pueden tener cualquier cantidad de bandas, además cada imagen puede contener metadatos que contengan información como la ubicación, el tiempo de recopilación y las condiciones en las que se recolectaron o procesaron las imágenes. Las imágenes de GEE están preprocesadas para permitir un acceso rápido y eficiente. (Gorelick et al., 2017). Cabe destacar que la plataforma permite a los usuarios descargar información en formato ráster o vectorial, además, a pesar de la posibilidad de procesar los datos en la nube, existe la función para poder descargar la información generada en formato GeoTIFF al almacenamiento Google Drive del usuario (Perilla y François, 2020).

El conjunto de datos de Sentinel-5P disponibles de GEE son índices de metano, dióxido de azufre, ozono, dióxido de nitrógeno, formaldehído, monóxido de carbono, nubes y aerosoles UV. Todas las colecciones de imágenes de Sentinel-5P excepto el metano, vienen dos versiones: Near Real Time (NRTI) y Offline (OFFL), el metano solo tiene esta última. Los datos NRTI cubren un área más pequeña que los OFFL, pero están disponibles en menos tiempo (Páez, 2022).

Los datos originales de Sentinel-5P son Nivel 2 (L2) y vienen agrupados por tiempo mas no por latitud y longitud, para hacer posible la ingesta de datos en GEE se debe convertir en datos L3 manteniendo una sola cuadrícula por órbita es decir no se realiza ninguna agregación entre productos, dicha conversión se realiza mediante la herramienta harpconvert donde los datos se filtran para eliminar píxeles con valores de control de calidad inferiores a: 80 % para AER_AI (Índice de aerosol UV), 75 % para la banda de columna vertical troposférica de NO_2 y 50 % para todos los demás conjuntos de datos excepto para O_3 y SO_2 (Google Developers s.f.).

4.2.7.1. Sentinel-5P OFFL NO_2: Dióxido de Nitrógeno Fuera de Línea.

NO_2 se usa para representar las concentraciones de los óxidos de nitrógeno en general, debido a que durante el día, en presencia de la luz solar, un ciclo fotoquímico que involucra ozono convierte NO en NO_2 y viceversa en minutos (Google Developers s.f.). La banda de NO_2 proporciona datos de columna vertical troposférica de NO_2 en mol/m^2. Estos datos están disponibles desde el 28 de junio del 2018 en una resolución de 1113,2 metros (Páez, 2022).

4.2.7.2. Sentinel-5P OFFL SO_2: Dióxido de Azufre Fuera de Línea

Este conjunto de datos proporciona imágenes fuera de línea de alta resolución de concentraciones de SO_2 atmosférico. Esta recopilación proporciona datos desde el 5 de diciembre de 2018 con una resolución de 1113,2 metros.
El conjunto de datos SO_2 entrega datos en mol/m_2. Para este producto L3 SO_2 , el índice de absorción de aerosol se calcula con un par de mediciones en las longitudes de onda de 340 y 380 nm. El producto L3_AER_AI tiene el índice de absorción de aerosol calculado utilizando las longitudes de onda de 354 nm y 388 nm (Google Developers s.f.).

4.2.8. Datos de Reanálisis ERA 5

ERA 5 es un conjunto de datos de alta precisión producido por el Servicio de Cambio Climático de Copérnico y el último reanálisis realizado por el Centro Europeo de Previsiones Meteorológicas a Medio Plazo (ECMWF) el reanálisis climático combina observaciones históricas y modelos para generar series temporales consistentes de múltiples variables climáticas proporcionan una descripción completa del clima observado a medida que ha evolucionado durante las últimas décadas (Ssenyunzi et al., 2020).

La plataforma proporciona datos por hora, día, semana y mes sobre diversas variables atmosféricas a diferentes presiones tanto de la superficie terrestre como del mar y engloba un periodo desde 1979 hasta el presente. ERA 5 se produce usando asimilación de datos

4D-Var y pronósticos modelo en CY41R2 del Sistema de Pronóstico Integrado (IFS) del ECMWF, con 137 niveles híbrido sigma/presión en la vertical y el nivel superior en 0,01 hPa. Los datos atmosféricos están disponibles en estos niveles y también se interpolan a 37 niveles, el mismo que tiene una resolución de 31 km. Entre las variables que dispone ERA 5 están: humedad relativa y específica, temperatura, componentes del viento dirección y velocidad, entre otras, las tres últimas variables fueron obtenidas para el presente estudio (ECMWF).

4.2.9. Datos de Geovanni NASA

Giovanni es una aplicación desarrollada por el Centro de Servicios de Información y Datos de Ciencias de la Tierra Goddard de la NASA (GES DISC) que permite a los investigadores explorar datos rápidamente y analizar de forma fácil y directa variaciones espacio-temporales, condiciones inusuales y patrones interesantes en línea antes de descargar los datos opcionalmente (Li et al., 2019). GES DISC brinda acceso a un conjunto como Land Data Assimilation System (FLDAS) (Jacob & NASA/GSFC/HSL, 2021), North American Land Data Assimilation System (NLDAS) (NLDAS project, 2021), Modern-Era Retrospective analysis for Research and Applications (MERRA) (GMAO, 2021), IMERG, GLDAS, entre otros, que permiten la obtención de datos tanto diarios, mensuales y cada tres horas para el posterior análisis y uso en investigaciones. Los formatos de descarga compatibles incluyen NetCDF, GeoTIFF y KMZ (EARTHDATA-NASA).

4.2.10. Datos meteorológicos de ERA 5 y Giovanni NASA

La contaminación del aire es causada por la presencia de gases y sustancias tóxicas, la misma que es afectada por los factores meteorológicos del lugar, como la temperatura, precipitación y velocidad del viento, entre otros (Le y Cha, 2018). Sin embargo (Qi et al., 2018) menciona que los factores que afectan el modelo de calidad del aire más que otros son: la velocidad del viento (del norte), la temperatura, la velocidad del viento (del este), radiación y la precipitación, en comparación con la presión barométrica y la humedad.

4.2.10.1. Precipitación

En meteorología, la precipitación es cualquier forma de lluvia que cae de la atmósfera y llega a la superficie terrestre. Esta variable afecta de forma directa a la concentración de contaminantes en la atmósfera debido a que en temporadas de sequía los datos de precipitación son bajos caso contrario cuando existe altos valores de precipitación los datos de concentración son menores (Muñoz et al., 2007).

4.2.10.2. Temperatura

La temperatura cumple una función importante sobre la calidad del aire debido a que cuando las temperaturas son bajas la calidad del aire es mejor mientras que la calidad del aire disminuye cuando aumenta la temperatura. Cabe recalcar que la temperatura varía durante el día, al amanecer las temperaturas son bajas y en el transcurso del día las temperaturas se elevan finalmente al llegar la noche la temperatura disminuye (Romero et al., 2010).

4.2.10.3. Radiación

La radiación es un proceso generado de forma natural para el cumplimiento de diversos procesos físicos y químicos. Sin embargo, en conjunto con el actual desbalance climático ocasionado por el efecto invernadero ocasiona problemas tanto individuales como interactivos en los sistemas biológicos. Además, la radiación también ha sido ligada en el aumento del calentamiento global a través de la basura vegetal y por la estimulación de gases de efecto invernadero por las plantas (Bornman et al., 2015).

4.2.10.4. Velocidad y Dirección del Viento

La velocidad del viento tiene un efecto significativo en la concentración y acumulación de contaminantes en el aire. El tipo de impacto contaminante en la zona urbana depende del

tráfico, su origen y el tiempo de residencia en la atmósfera. Los vientos están relacionados con la dinámica horizontal de la atmósfera y en base a estos es posible conocer la dirección de desplazamiento del contaminante, la velocidad de dispersión y la turbulencia. Los vientos locales mueven el aire desde zonas de alta presión a baja presión y determinan los vientos dominantes de un área (García et al., 2014). Esto provoca el arrastre de contaminantes de una zona a otra o la permanencia del mismo en un lugar establecido.

4.2.10.4.1. Componente U del Viento

Este parámetro es el componente este del viento. Es la velocidad horizontal del aire que se mueve hacia el este, un signo negativo indica que el aire se mueve hacia el oeste (ECMWF).

4.2.10.4.2. Componente V del Viento

Este parámetro es el componente norte del viento. Es la velocidad horizontal del aire que se mueve hacia el norte, un signo negativo indica aire moviéndose hacia el sur (ECMWF).

4.2.11. Inteligencia artificial - Machine Learning

Durante los últimos años, las técnicas basadas en machine Learning han abordado muchas áreas de la industria, incluida áreas de la medicina, procesos de fabricación, recolección de datos y hasta conducción autónoma. Esta ciencia disruptiva tiene como objetivo reconocer patrones y de esta manera tratar problemas imperceptibles. Como resultado de los avances se ha demostrado que los métodos de aprendizaje automático pueden ser utilizados en tareas específicas (Carleo et al., 2019). Las técnicas de análisis de datos nacen a mediados del siglo XX a medida que la tecnología computacional avanzaba a pasos agigantados. Y esto ha permitido la creación de nuevas técnicas de análisis de datos como la regresión y las redes neuronales artificiales, las cuales aparecieron en los años de 1950. Sin embargo, algunos métodos de aprendizaje profundo y los métodos de entrenamiento aparecieron décadas después abriendo paso al tratamiento de datos

conjuntos en porciones elevadas y de esta manera demostrando que este tipo de método es capaz de identificar patrones complejos o imperceptibles para el ojo humano (Biamonte et al., 2017).

4.2.12. Redes Neuronales

Una red neuronal puede definirse como una máquina diseñada originalmente para imitar la forma en que el sistema nervioso de un ser vivo realiza una determinada tarea. Para lograr esto la neurona está formada por un conjunto de unidades de procesamiento que se encuentran conectadas entre sí llamadas neuronas.

Cada neurona recibe como entrada un conjunto de señales las cuales las pondera e integra y transmite el resultado a las neuronas conectadas a ella. Todas las neuronas se encuentran conectadas entre sí mediante un peso. En los pesos se suele guardar la mayor parte del conocimiento que la red neuronal tiene sobre la tarea en cuestión. El proceso mediante el cual los pesos se ajustan con la finalidad de lograr un determinado objetivo se denomina aprendizaje o entrenamiento y el procedimiento utilizado para esta acción se conoce como algoritmo de aprendizaje o algoritmo de entrenamiento (Pérez, 2002).

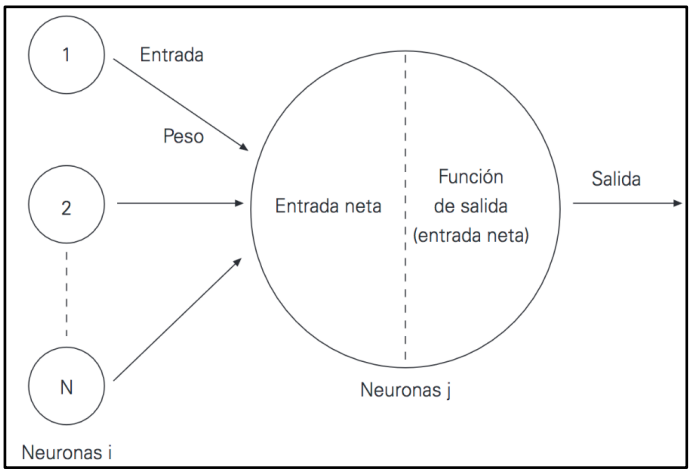

Gráfico 2. Funcionamiento General de una neurona artificial (10 Introducción a Las Redes Neuronales Artificiales | Introducción al Software Estadístico R)

4.2.12.1. Redes Neuronales Recurrentes

Las redes neuronales recurrentes (RNN) nacen de las redes neuronales artificiales en donde como principal diferencia estas poseen conexiones recurrentes, permitiendo de esta manera trabajar con datos secuenciales para el reconocimiento y la predicción de patrones. Las RNN están formadas por estados ocultos de alta dimensión los cuales funcionan como la memoria de la red y están condicionados con el estado de la capa anterior, permitiendo que estas redes almacenen, recuerden y procesen datos durante largos periodos de tiempo (Salehinejad et al., 2018).

Una de las características más importantes de las RNN es compartir sus parámetros. Sin compartir parámetros, el modelo establecerá parámetros únicos para representar a cada dato en una secuencia y, por lo que la inferencia no es posible sobre secuencias de longitud variable. El impacto de esta limitación es evidente en el procesamiento del lenguaje natural. Las redes multicapa tradicionales puede fallar porque generarán explicaciones lingüísticas para los parámetros establecidos para cada posición en una frase, Las RNN sin embargo, serían más adecuadas para la tarea ya que comparten pesos entre los datos espaciados secuencialmente en el ejemplo del lenguaje, las palabras de nuestra frase (Ariana, 2021).

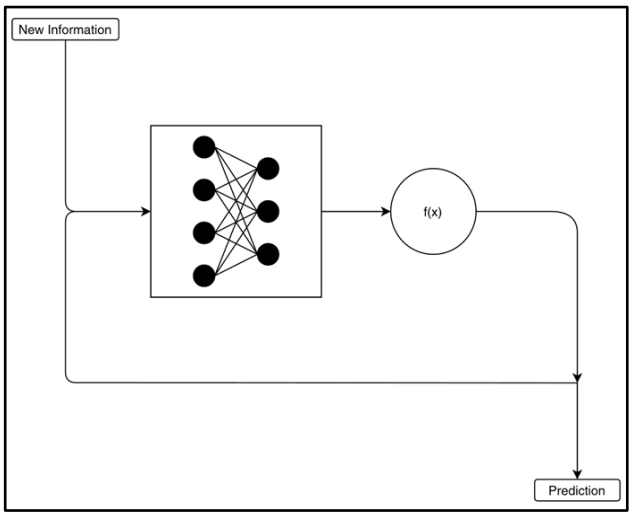

Gráfico 3. Diagrama de grado cíclico de una Red Neuronal Recurrente. (Ariana, 2021)

Las RNN generalmente aumentan la arquitectura de la red multicapa convencional con la adición de ciclos que conectan nodos adyacentes o pasos de tiempo. Estos ciclos

constituyen la memoria interna de la red que se utiliza para evaluar las propiedades del dato actual con respecto a los datos del pasado inmediato (Ariana, 2021).

Estas redes son utilizadas para una variedad muy grande de propósitos como la música, el texto y los datos de captura de movimiento (Graves, 2013). Además de que existen muchas tareas las cuales para su resolución se requiere que los datos sean tratados secuencialmente. Para la síntesis de voz, generación de música, generación de subtítulos es necesario que el modelo produzca salidas en forma de secuencia. Y para el análisis de videos, para la predicción de series temporales y la recuperación de información musical los modelos deben utilizar series secuenciales para aprender (Lipton et al., 2015).

4.2.12.2. Sobreajuste

El sobreajuste ocurre cuando un algoritmo reduce su error memorizando los datos de entrenamiento en lugar de aprender la verdadera relación que existe entre las variables analizadas. Debido a que es probable que el conjunto de datos que utilizamos para entrenar la red contengan ruido, errores o que los datos que representan nuestra línea base no son representativos con la realidad del problema al tener pocos datos nos encontraremos con que, al llegar a cierto punto, al seguir mejorando el resultado, la red en el conjunto de entrenamiento comenzará a degradar el rendimiento en el conjunto de prueba, en ese momento se empezará a producirse el sobreajuste (Galvan, 2021).

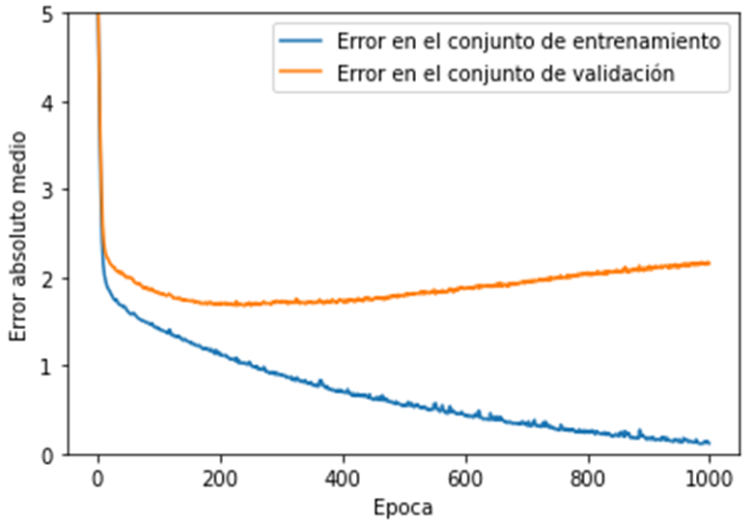

Gráfico 4. Ejemplo de cómo se produce el sobreajuste al entrenar una red. Elaborado por: Autores, 2022.

4.2.13. Random Forest

Los bosques aleatorios o Random Forest son un esquema propuesto por Leo Breiman en la década de 2000 para construir un conjunto de predictores con un conjunto de árboles de decisión que crecen en subespacios de datos seleccionados al azar (Biau, 2012). El reconocimiento del poder adquirido por el método de bosques aleatorios se debe a que no se limita con una pequeña gama de procesos sino que este puede ser aplicado a varios problemas de predicción y además de que los parámetros que se aplican no son variados. Un árbol de regresión es un árbol binario que se compone de un nodo raíz, de nodos internos y de nodos terminales llamados hoja, representado en círculos como se representa en el gráfico 5. Cada nodo interno representa un subconjunto de las observaciones y un test binario que resulta en la generación de dos nodos hijos. El algoritmo de regresión consiste en dividir el espacio de las variables X en particiones homogéneas para obtener en la salida del árbol una predicción de la variable en función de los valores de X (Chardon, 2019).

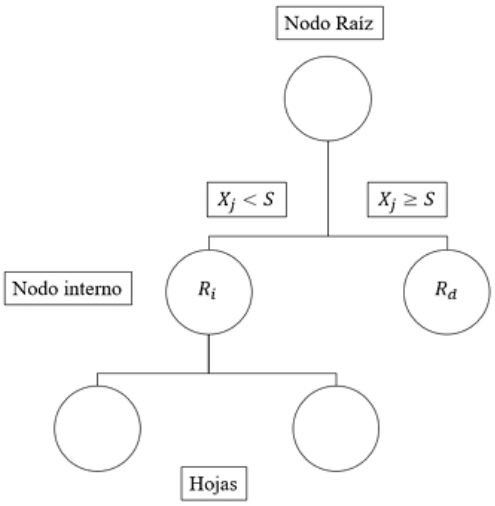

Gráfico 5. Representación gráfica de un árbol de regresión. (Ghardon, 2019).

4.2.13.1. Arboles de Decisión

Los árboles de decisión con métodos comúnmente utilizados en tareas de clasificación y regresión, la idea vino de un árbol ordinario. Su estructura se compone de una raíz y nodos,

ramas y hojas, de la misma manera se construye un árbol de decisión, de nodos que representan círculos y las ramas son representadas por segmentos que conectan los nodos. El árbol comienza hacia abajo, el nodo donde comienza se llama nodo raíz, el nodo de los extremos se llama nodo hoja. Dos o más ramas se pueden extender desde cada nodo interno. Un nodo representa una determinada característica mientras que las ramas representan un rango de valores. Estos rangos actúan como puntos de partición para el conjunto de valores de la característica dada (Ali et al., 2012).

4.2.13.2. Construcción de un Bosque aleatorio

La construcción de los bosques aleatorios está basada en la combinación de múltiples árboles de decisión. Se utiliza el algoritmo de partición binaria recursiva CART para la construcción de estos árboles ya que tiene la ventaja de poder utilizar tanto variables categóricas como numéricas como regresoras y acepta variables de estudio de ambas tipologías. Los principales parámetros a considerar para la construcción del modelo se presentan en el siguiente cuadro.

Tabla 1. Parámetros para construir un Bosque Aleatorio.

Parámetro	*Definición*	*Valor por defecto*
mtry o m	Número de variables aleatorias usadas en cada partición	p/3 en regresión
n	Número de observaciones	N
max_depth	Profundidad máxima del árbol	Ninguno
replace	Usar o no reemplazo en la muestra	SI
min.node.size	Número mínimo de unidades en un nodo final	1 clasificación y 5 regresiones
num.trees	Número de árboles en el RF	500 o 1000
splitrule	Criterio usado para hacer las divisiones	Gini impurity, Variance

Para la construcción de los bosques aleatorios nos basamos en el Algoritmo de Breiman (Cutler et al., 2012). A continuación, con los parámetros definidos seleccionamos el conjunto de entrenamiento S para cada uno de los *j* árboles que comprendan en RF.

1. Obtener una muestra Bootstrap de tamaño *n* de *S* obteniendo así *Sj*
2. Usar *Sj* como conjunto de entrenamiento aplicar partición binaria recursiva (CART) pero con las siguientes especificaciones:
   I. Empezar con todas las observaciones en el nodo raíz
   II. Para cada partición hasta que se cumpla el criterio de parada realizar:
      1) Seleccionar *mtry* predictores aleatorios.
      2) Hacer la partición óptima (según el *splitrule*) sobre esos *mtry* predictores.

Las predicciones para regresión serán la medida de las observaciones que haya en el nodo terminal correspondiente y la moda para los árboles de clasificación

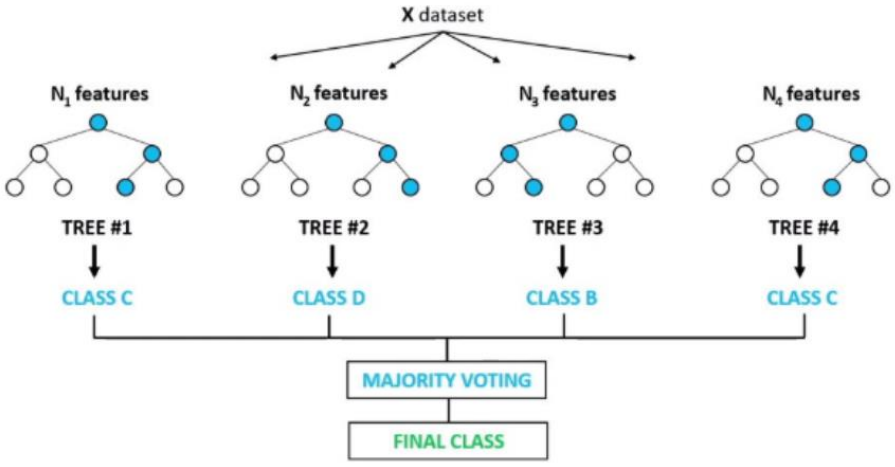

Gráfico 6. Diagrama de funcionamiento de Random Forest. (Cardellino, 2021).

Una de las mayores ventajas que posee es que es fácilmente paralelizable y por lo tanto se acopla a varios sistemas de la vida real. Puede ser aplicado con éxito en ciencias de la predicción de la calidad del aire, en la ecología, para el reconocimiento de objetos 3D y bioinformática (Biau y Scornet, 2016). Este método es utilizado para la clasificación y regresión de datos, el cual está conformado por una serie de árboles de decisión aleatorios formados en la fase de entrenamiento el cual predice resultados promediados. Y a pesar

de que son ampliamente utilizados debido a su alto rendimiento se desconoce en su totalidad las propiedades matemáticas aplicadas, esto debido a que existe una disparidad entre la teoría y la práctica por la dificultad que representa analizar simultáneamente la estructura del árbol y el proceso de aleatorización (Scornet et al., 2015).

4.2.14. Lenguaje de programación Python

Python es un lenguaje de programación creado por Guido van Rossum a inicios de los 90´s (Ortiz, 2010). Se trata de un lenguaje interpretado o de script, con tipado dinámico donde no hay necesidad de declarar el tipo de dato de cada variable, fuertemente tipado, donde no se permite tratar a una variable como si fuera de un tipo diferente al que es, multiplataforma y orientado a objetos donde los conceptos reales relevantes para nuestro problema se transforman en clases y objetos en nuestro programa (González, 2011).
Un lenguaje interpretado o de script es aquel que se ejecuta utilizando un programa intermedio llamado intérprete, en lugar de compilar el código de lenguaje de máquina compilado que una computadora puede entender y ejecutar directamente. Los lenguajes compilados tienen la ventaja de ser más rápidos, sin embargo, los lenguajes interpretados son más flexibles y más portátiles (González, 2011). El hecho es que las tecnologías libres y abiertas tienen ventajas significativas sobre las tecnologías propietarias. La más importante de ellas es que puede usarse sin ningún costo de licencia (Ortiz, 2010).

Este lenguaje de programación en las últimas décadas está ganando popularidad entre los intérpretes científicos y los desarrolladores, debido a que existe una gran diferencia con el lenguaje de programación R el cual simplemente está destinado al análisis de datos estadísticos. El lenguaje Python es un lenguaje muy variado y es utilizado para un sin número de aplicaciones distintas, como para el desarrollo de sitios web, desarrollo de software y juegos, computación científica y el acceso a bases de datos (Hao y Ho, 2019).

5. METODOLOGÍA

5.1. Área de Estudio

Para determinar las áreas de estudio, se seleccionaron las tres ciudades más importantes que tengan alta densidad poblacional, altas concentraciones de contaminantes atmosféricos, reconocimiento y estén distribuidas de forma idónea para el estudio. Las ciudades elegidas para representar al Ecuador fueron: Quito, Guayaquil y Cuenca.

De esta manera, se eligió la ciudad de Quito porque ha sido el referente de la gestión a nivel nacional y al mismo tiempo, se considera un área nacional propensa a la contaminación del aire (Páez, 2022). Para el caso de la ciudad de Cuenca, se evidencia que la contaminación del aire ha alcanzado niveles significativos en los últimos años (Moyano, 2017), esto debido a muchos factores como el crecimiento poblacional y el aumento de la industria, además, teniendo en cuenta de que es una ciudad de gran índole turístico, por este motivo se ha tomado como área de estudio para el presente trabajo. Finalmente, se ha elegido a Guayaquil ya que es la ciudad más grande del país con una población de 1 952.029 habitantes, su economía gira entorno a la agricultura, comercio internacional a través de sus puertos y la movilidad y transporte que como consecuencia se presentan emisiones atmosféricas elevadas (Zambrano, 2014).

A continuación, en el Gráfico 7 se presentan gráficamente las áreas de estudio detalladas anteriormente.

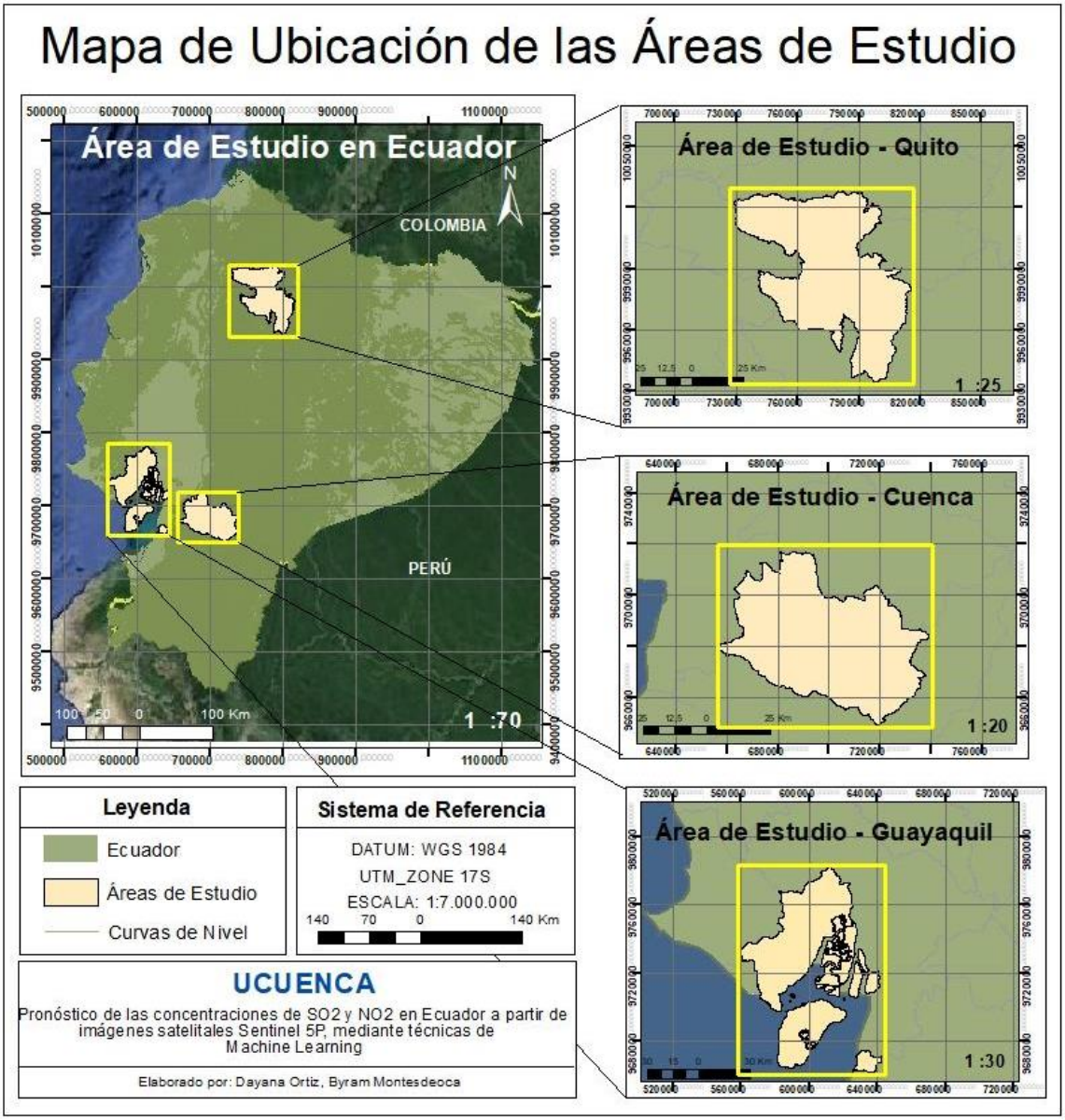

Gráfico 7. Mapa de la Zona de Estudio: Ecuador. Elaborado por: Autores, 2022.

5.2. Obtención de Imágenes satelitales

En esta investigación se utilizaron datos de concentración de NO_2 Y SO_2 a partir de imágenes satelitales Sentinel-5P mediante Google Earth Engine. Por otro lado las variables meteorológicas de precipitación, radiación, temperatura y componente U y V del viento

fueron obtenidas de dos plataformas distintas Giovanni NASA Y ERA 5 todas en un periodo de dos años desde el 1 de enero de 2019 hasta el 31 de diciembre de 2020.

5.2.1. Imágenes de Sentinel-5P

Para obtener las imágenes diarias de las concentraciones de los gases NO_2 Y SO_2 se empleó un código de programación utilizando algunas de las funciones y comandos de la interfaz de programación de Aplicaciones API de la plataforma GEE, escritas en lenguaje de programación JAVA de código abierto (Perilla y François, 2020). El proceso del mismo se detalla a continuación

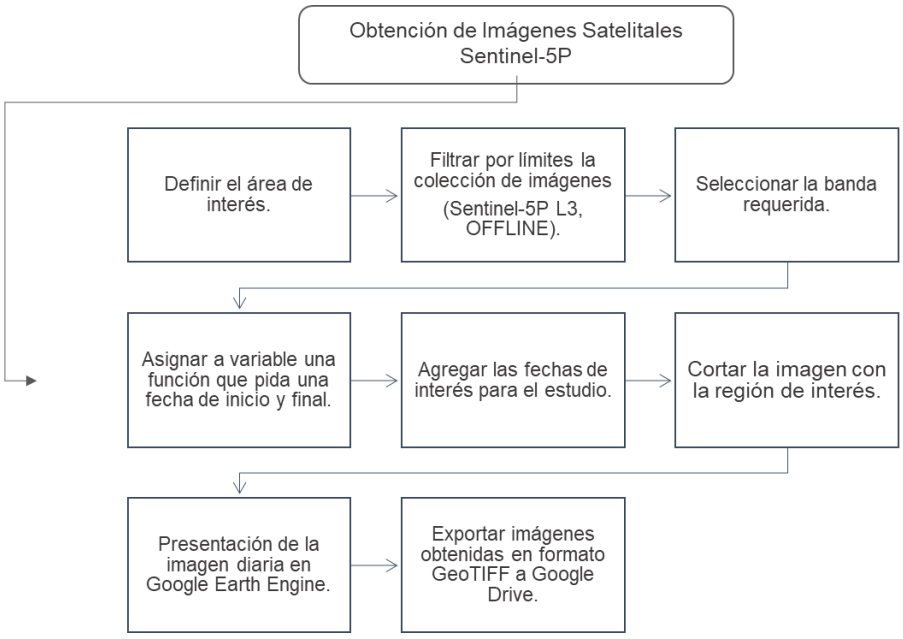

Gráfico 8. Diagrama de Flujo de proceso de obtención de imágenes satelitales Sentinel-5P. Elaborado por: Autores, 2022.

A continuación se detallan las características de las variables de concentración de contaminantes de dióxido de nitrógeno y dióxido de azufre las mismas que fueron descargadas de la plataforma de GEE.

Tabla 2. Descripción de las variables de concentración de contaminantes. Elaborado por: Autores, 2022.

Plataforma	***Variable***	***Descripción***	***Resolución***	***Unidades***
Satélite Sentinel 5P a través de la plataforma Google Earth Engine	Dióxido de Nitrógeno NO_2	Columna vertical troposférica de NO_2	0,01 x 0,01 grados	umol/m^2
	Dióxido de Azufre SO_2	Densidad de columna vertical de SO_2 a nivel del suelo	0,01 x 0,01 grados	umol/m^2

5.2.2. Imágenes de datos meteorológicos.

Para la estimación de la concentración espacial de contaminantes se optó por utilizar de la misma manera datos meteorológicos recuperados de imágenes satelitales. Esto debido a que son factores directos los cuales afectan de varias maneras la concentración de los contaminantes en las zonas de análisis.

Para este análisis se utilizaron dos plataformas para la obtención de las imágenes satelitales. Se utilizó la plataforma ERA 5 para obtener los datos de temperatura y los componentes U y V del viento. Para obtener los datos de radiación y precipitación se utilizó la plataforma Geovanni de la NASA. La plataforma ERA permite descargar los datos en formato NetCDF y los datos se pueden obtener en una escala temporal horaria por lo que al momento de utilizarlos para el método se tuvo que promediar y de esta manera los datos podían ser transformados a una escala temporal diaria. Por otra parte, la plataforma Geovanni de la NASA permite descargar datos a una escala temporal variada por lo tanto debido al objetivo del proyecto los datos fueron descargados diariamente y en formato NetCDF. Todas las imágenes descargadas se encontraban en el Datum 84 (World Geodetic System 1984) y en el sistema de coordenadas UTM (Universal Transverse Mercator)

A continuación, en el gráfico 9 se especifica el proceso de manera más detallada del cómo se obtuvieron las diferentes variables meteorológicas utilizando las dos plataformas.

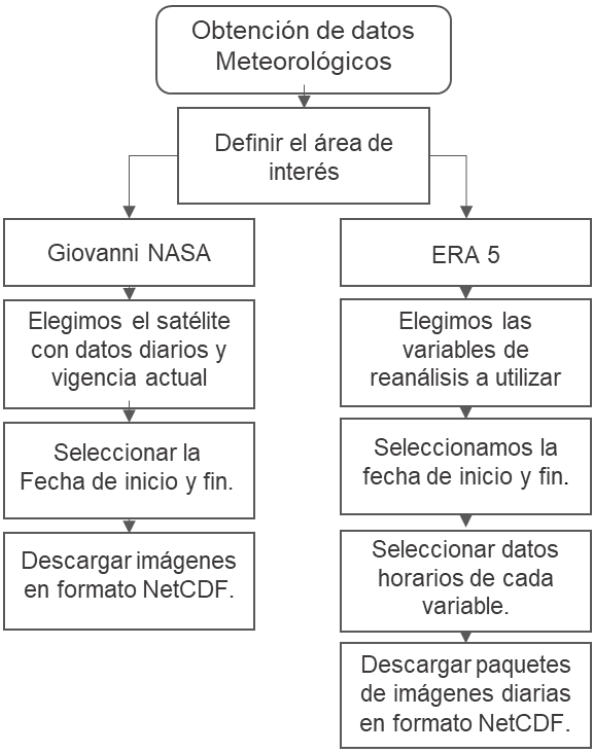

Gráfico 9. Diagrama de flujo de obtención de Imágenes satelitales de variables meteorológicas. Elaborado por: Autores,2022.

Además, en la tabla 3 se especifica los diferentes parámetros que se descargaron en conjunto con las variables ya descritas.

Tabla 3. Descripción de las variables meteorológicas. Elaborado por: Autores, 2022.

Plataforma	***Variable***	***Descripción***	***Resolución***	***Unidades***
Giovanni NASA	Precipitación	Es el proceso en el que el vapor de agua se condensa y forma gota de lluvia	0,1 x 0,1 grados	mm/día
	Radiación	La radiación de onda corta es la cantidad de radiación solar incidente absorbida en la superficie terrestre por área.	0,1 x 0,1 grados	Wm-2
ERA 5	Temperatura	Es la temperatura en la atmósfera	0,25 x 0,25 grados	K
	Componente U del viento	Es la velocidad horizontal del aire que se mueve hacia el este.	0,25 x 0,25 grados	m/s
	Componente V del viento	Es la velocidad horizontal del aire que se mueve hacia el norte.	0,25 x 0,25 grados	m/s

5.3. Procesamiento de Imágenes Satelitales

5.3.1. Extracción de los datos de las concentraciones de contaminantes de las Imágenes Satelitales Sentinel-5P.

La extracción de los datos se realizó a través del software libre y de código abierto QGIS 3.16.2. Mediante la herramienta “Píxeles ráster a puntos” y “Agregar atributos de geometría” la primera permitió convertir los valores de los píxeles de las capas ráster en capas vectoriales, la segunda nos posibilitó brindar coordenadas de latitud y longitud a los datos extraídos y finalmente extraer los puntos a archivos csv.

5.3.2. Extracción de los datos de las variables meteorológicas a partir de Giovanni NASA Y ERA 5.

De manera similar a los datos de concentración anteriores, los datos meteorológicos fueron procesados en el software QGIS 3.16.2, con la diferencia que para ellos se utilizó la herramienta “*Resample*” con el fin de igualar el tamaño del píxel de las imágenes Sentinel 5P mediante el método de interpolación bilineal. Este método de interpolación bilineal reemplaza cada píxel faltante con un promedio ponderado de los píxeles más cercanos en el límite de los 4 píxeles adyacentes, los valores utilizados son inversamente proporcionales a la distancia entre el píxel de origen y los píxeles de destino (Agrafiotis, 2014). Para el caso de los datos obtenidos a partir de ERA 5 se tuvo que promediar los datos horarios para convertirlos en datos diarios.

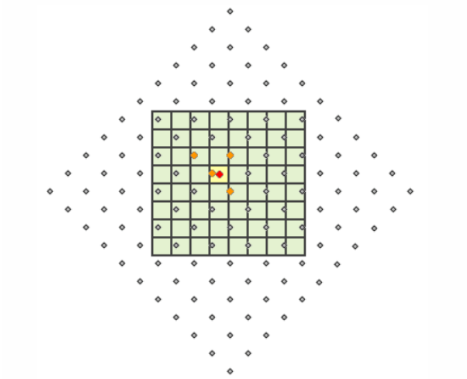

Gráfico 10. Método de Interpolación Lineal. (Esri, 2019).

Una vez los datos de las imágenes satelitales se hayan remuestreado, se procedió a utilizar las herramientas ya mencionadas para extraer los datos en formato Excel.

5.3.3. Limpieza de datos

Después de extraer los datos se procedió a limpiar los valores que debido al ruido se observan valores negativos en particular son regiones limpias excepto para valores atípicos es decir para columnas verticales menores a -0,001moles/m^2. Además, se utilizó la herramienta *Boxplot* para eliminar los valores atípicos extremos, esta herramienta es un método de visualización del grado de dispersión de los datos donde se elige cuartiles y el rango intercuartílico como bases de juicio, que no se verá afectado por los valores atípicos (Li et al., 2016)

Posteriormente los datos limpios fueron rellenados mediante el método de Imputación K-Nearest Neighbors (KNN) el cual almacena todas las observaciones disponibles y clasifica nuevos casos basado en una media semejante (Bello et al., 2019). Posteriormente, se calcularon los promedios por áreas de estudio con las concentraciones diarias correspondientes, esto se repitió para cada área de estudio.

Una vez extraídos todos los campos necesarios tanto las variables dependientes NO_2 y SO_2 y las variables independientes Precipitación, Radiación, Temperatura, Componente U y V del viento se estableció una base de datos para comenzar a trabajar con los métodos establecidos.

5.4. Determinación de las concentraciones de NO_2 y SO_2 mediante Redes Neuronales Recurrentes

Para diseñar una red neuronal para predicciones se ha propuesto 8 pasos, los cuales los tres primeros se basan en la obtención y preprocesamiento de los datos, por lo tanto partimos del paso número 4 el cual consiste en la normalización y definición del conjunto de entrenamiento, prueba y validación, en el quinto paso se define paradigmas de las redes neuronales artificiales o la arquitectura, el paso 6 se basa en el entrenamiento de la red neuronal donde se especifica el número de iteraciones de entrenamiento y la bondad de ajuste, después, definimos los criterios de evaluación del modelo y finalmente la implementación de la red (Calopiña, 2014).

5.4.1. Normalización

Debido a las variables que se tomarán en cuenta para la realización del caso de estudio y la diferencia de escalas que presenta cada una, se optó por realizar una normalización para de esta manera facilitar el entendimiento de los resultados obtenidos, esto debido a que existe una gran variabilidad de escalas y rangos en los datos.

Para realizar la normalización del datum se optó por utilizar el método de Distribución Normal Estándar, la cual es normalmente el método más utilizado y esto permite que los datos se aproximen a una distribución normal estándar. Es decir, los datos se acercan a una media aritmética de un valor correspondiente a cero y una desviación estándar de un valor de uno (Salazar et al., 2018). A continuación, en la ecuación 1 se describe la ecuación de la Distribución Normal Estándar.

Ecuación 1. Distribución Normal Estándar.

$$y = \frac{x - \mu}{\Theta}$$

Donde "x" representa el valor que se desea normalizar, "μ" representa la media aritmética de la variable que se desea estandarizar y "Θ" corresponde a la desviación estándar.

5.4.2. Arquitectura

Hablamos de arquitectura cuando nos referimos al número de neuronas y la manera en la que se conectan en una red neuronal. Estas neuronas se agrupan en unidades estructurales denominadas capas las cuales al conectarse se le adiciona un valor llamado peso el cual determina la intensidad con la que el valor de la neurona es asociado con la siguiente neurona.

Existen tres tipos de capas: Las de entrada, las ocultas o intermedias y las de salida. La determinación del número de capas y el número de neuronas en la capa oculta es un proceso complejo debido a que esto depende netamente del objetivo a realizar y la complejidad del mismo. En ese caso si lo que se desea obtener es mayor precisión en los resultados los más óptimo sería utilizar una gran cantidad de capas ocultas, sin embargo si un impedimento es el tiempo se puede optar por utilizar un número bajo de capas ocultas. Por otro lado, es muy importante que antes del diseño del método se analicen los datos para determinar una aproximación óptima del número de capas y número de neuronas para de esta manera evitar problemas de sobreajuste (Uzair & Jamil, 2020).

Ya que toda red neuronal posee una capa de entrada para la realización se utilizó el número de neuronas correspondiente a todas las variables independientes. Para el número de capas ocultas no existe una fórmula que ayude a determinar su número óptimo, sin embargo, debido a la problemática y basándonos en el teorema de kolmogorov se optó por utilizar una capa oculta dado que se ha demostrado que para problemas de la misma magnitud es suficiente utilizar una capa (Stathakis, 2009). Y para el número de neuronas ocultas se utilizó 16, 32 y 64, ya que para determinar el número óptimo de neuronas se realiza es un proceso de prueba error y de esta manera evitar además un sobre ajuste con los datos de entrenamiento.

5.4.3. Función de Activación

La función de activación es un proceso que distorsiona de cierta manera los valores transmitidos por la neurona a la siguiente neurona añadiendo deformaciones no lineales y facilitando que la neurona realice un aprendizaje progresivo.

Se optó por utilizar la función de activación llamada Rectified Linear Unit (Unidad Lineal Rectificada) abreviada como ReLU, la misma que se comporta como una función lineal cuando es positiva y constante a 0 cuando el valor de entrada es negativo. Y se define de la siguiente manera (Banerjee et al., 2019).

Ecuación 2. Función de activación Rectified Linear Unit (RELU).

$$f(x) = máx\ (0, x)$$

5.4.4. Algoritmo de Optimización

Los algoritmos de optimización se describen como suavizadores de función numérica los cuales son utilizados para reducir el error. Estos optimizadores ayudan a reducir la pérdida que sufre el modelo cuando este se está entrenando y se utiliza para determinar el grupo de datos destinados al pronóstico los cuales se accede en el modelo mismo (Vani y Rao, 2019).

Para el algoritmo de optimización se utilizaron tres diferentes optimizadores de la librería keras entre ellos el optimizador RMSprop utiliza el teorema de momentum que devuelve una lista de tres valores: el recuento de iteraciones, seguido del valor de la raíz cuadrada del núcleo y el sesgo de la capa densa única (Hinton et al., 2012), el algoritmo SGD es el optimizador de gradiente estocástico el cual incluye compatibilidad con el impulso, la disminución de la tasa de aprendizaje y el impulso de Nesterov (Velumani et al., 2021) y *Optimizer.Adam* donde implementa el algoritmo de Adam es un método de descenso de gradiente estocástico que se basa en la estimación adaptativa de momentos de primer y segundo orden (Kigman y Lei Ba, 2017). El último fue establecido como optimizador para nuestra red ya que mostró mejores resultados.

5.4.5. Entrenamiento y validación

Para crear el modelo de entrenamiento y realizar el ajuste se utilizó la métrica de Mean Square Error MSE ya que es la ideal para entrenar los modelos de regresión ya que toma el error entre un valor estimado y el valor real, lo eleva al cuadrado y calcula su promedio

(Lavín, 2021). Para efectividad del método lo que se realizará es una división de los datos recopilados 80-20, en donde 80 % de los datos serán utilizados para entrenar el método, es decir para determinar el valor de los pesos y el 20 % se utilizará para validar el método, en donde lo que se desea es medir el error. A esto se lo denomina validación cruzada. Los pesos se calculan de manera iterativa, de acuerdo a los valores de entrenamiento con el objetivo de reducir el error en la salida obtenida por la red neuronal y la salida deseada (Camilo, 2019). Con la finalidad de tener un significado estadístico se determinó el número de iteraciones o el número de veces que el modelo repetiría el entrenamiento formando distintas agrupaciones con un valor de 30. El número de Épocas es el número de veces que se debe repetir el modelo observando la base de datos con la finalidad de hacer los ajustes en los pesos de la red, para bases de datos con pocos datos las épocas pueden variar entre 500 y 1 000 para una base de datos mayor el número de épocas aumentará respectivamente (Kumar, 2020). En cuanto a la validación del modelo se determinó usar el 20 % de los datos utilizados para entrenar el modelo que corresponde al 80 % del total de datos, el cual sirvió para monitorear el entrenamiento.

5.5. Determinación de las concentraciones de NO_2 y SO_2 mediante Random Forest

Como se ha comentado antes un bosque aleatorio (Random Forest) es una clasificación de conjunto que utiliza muchos árboles para predecir el resultado final del problema específico (Breiman, 2001).

En primer lugar, obtenemos el conjunto de datos tanto de los datos de concentración como las variables meteorológicas a analizar el cual no se necesita ser normalizado anteriormente debido a que no causó ningún impacto en los resultados. Luego se seleccionó una muestra aleatoria del conjunto de datos y se construyó un árbol decisión para cada muestra que calculamos. Para ellos se debe considerar varias características entre ellas las más importantes son el número de árboles el cual para el tamaño del conjunto de datos podrían considerarse 100, 500 y 1 000; la profundidad máxima el cual son los niveles a los cuales se va a ir derivando el árbol en cada uno de los árboles de decisión que se va a generar y el mínimo de números de muestras para que sea

considerado un nodo hoja donde se considera cuántos elementos va a quedar al final de la ramificación (Geurts et al., 2006) .

Para el caso de estudio se realizaron dos tipos de bosques, en el primero se consideró la profundidad máxima de las raíces, las cuales se determinaron a través del método de prueba error, se consideró que 5, 15 y 25 son los valores más adecuados para dicha característica y el segundo se tomó como prioridad el número mínimo de muestras para que se considere un nodo hoja considerando la misma cantidad de 5, 15 y 25. El número de repeticiones se consideró el mismo que para el método anterior el cual se determinó 30 repeticiones.

Para activar el logaritmo de Random Forest utilizamos la función de *RandomForestRegressor* ya que utiliza el promedio para mejorar la precisión predictiva y controlar el sobreajuste.

5.6. Análisis Estadístico de los Datos de Predicción

En este estudio se realizó la comparación de las predicciones entre áreas de estudio y la comparación de las predicciones entre los días 1 hasta el día 5 hacia adelante. Para realizar las pruebas estadísticas a los datos de la investigación se realizó en el mismo código de programación Python donde se realizó la predicción. Los análisis estadísticos que se realizaron para ambos métodos de predicción son los mismos.

5.6.1. Determinación de la bondad de ajuste

Las pruebas de bondad de ajuste se utilizan para contrastar si los datos de la muestra pueden considerarse que proceden de un determinado conjunto de observaciones, en resumen, permiten verificar que tipo de distribución siguen nuestros datos y por lo tanto qué pruebas podemos llevar a cabo en el contraste estadístico, para este caso se usó las métricas con los datos predichos con los datos del conjunto de entrenamiento. (Romero, 2016).

Para la predicción mediante Redes Neuronales Recurrentes ello se utilizó el coeficiente de determinación (R^2), el mismo que es una medida representativa para el cambio de porcentaje de una variable dependiente o explicativa la cual es explicada por un conjunto de variables independientes (Rodríguez et al., 2022). El coeficiente de determinación básicamente es el grado de variación que tienen 2 conjuntos de números al mismo tiempo y estos valores son más fáciles de entender para la mayoría de las personas y es básicamente el coeficiente de correlación elevado al cuadrado (Brown, 2003).

La ecuación es definida como:

Ecuación 3. Coeficiente de determinación.

$$r^2 = \left(\frac{\sum_{i=1}^{n} (y_i^{obs} - \bar{y}^{obs})\left(y_i^{sim} - \bar{y}^{sim}\right)^2}{\sqrt{\sum_{i=1}^{n} (y_i^{obs} - \bar{y}^{obs})^2} \sqrt{\sum_{i=1}^{n} (y_i^{sim} - \bar{y}^{sim})^2}} \right)^2$$

Donde y_i^{obs} es el rendimiento observado, y_i^{sim} es el rendimiento simulado, $\bar{y}^{obs}$ corresponde al rendimiento medio observado, $\bar{y}^{sim}$ es el rendimiento medio simulado y *n* es el número total de observaciones.

5.6.2. Determinación de pruebas de error

Las pruebas de error se usan para interpretar los datos obtenidos de nuestra predicción con el objetivo de saber si se ha cumplido o no los objetivos propuestos y la viabilidad de poder duplicar la metodología para otros fines. En este caso, para determinar el error se utilizaron las predicciones que da el modelo en el conjunto de test.

5.6.2.1. Error Cuadrático Medio

El Error cuadrático medio (RMSE) es la raíz cuadrada del promedio de la diferencia al cuadrado entre los valores obtenidos y las observaciones. Es utilizado como una métrica estadística estándar para medir el rendimiento del modelo en estudios de meteorología, calidad del aire e investigación climática. La suposición subyacente al presentar RMSE es que los errores no están sesgados y siguen una distribución normal por lo tanto usar RMSE

ayuda a brindar una imagen completa de la distribución del error (Chai y Draxler, 2014). El RMSE es definido como:

Ecuación 4. Error cuadrático medio.

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}$$

Donde *m* es la longitud de la serie temporal, y_i es el valor observado $\hat{y}_i$ es el valor simulado.

5.6.2.2. Sesgo Porcentual (PBIAS)

PBIAS proporciona información sobre la tendencia del modelo a sobreestimar (valores negativos) o subestimar (valores positivos) la variable de interés (Gupta et al., 1999). El valor óptimo de PBIAS es 0, lo que indica valores de baja magnitud de simulación precisa del modelo. Los valores positivos indican sesgo de sobreestimación del modelo y los valores negativos indican sesgo de subestimación del modelo (Moriasi et al., 2007).

PBIAS se calcula mediante la ecuación:

Ecuación 5. Sesgo porcentual.

$$\mathrm{PBIAS} = \left[\frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim}) * (100)}{\sum_{i=1}^{n}(Y_i^{obs})}\right]$$

Donde PBIAS es la desviación de los datos evaluados (%), Yi^{obs} y Yi^{sim} corresponden al rendimiento observado y simulado por el modelo respectivamente y *n* se refiere al número total de observaciones.

5.6.2.3. Error Porcentual Absoluto Medio (MAPE)

El MAPE es un indicador del desempeño que mide el tamaño del error en términos porcentuales. Se calcula como el promedio de los errores porcentuales sin considerar el signo. La métrica tiene valiosas propiedades estadísticas porque utiliza todas las observaciones y tiene la variación más pequeña de muestra a muestra (Niño, 2018).

Ecuación 6. Error porcentual absoluto medio.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{D_t - F_t}{D_t}\right|$$

Donde *n* es el número de muestras, *Dt* es el valor actual y *Ft* es la estimación. La función de pérdida de la medida es la del error absoluto.

6. RESULTADOS

En la investigación realizada durante el desarrollo de esta tesis no se cuenta con información de tráfico ni densidad de población, la base de datos utilizada únicamente contiene información sobre variables de concentración de contaminantes SO_2 y NO_2 y variables meteorológicas en estudios realizados se determinó que la variables meteorológicas que intervienen más en la concentración futura de los contaminantes son: Precipitación, Radiación, Temperatura, Velocidad y dirección del viento (Componente U y V del viento). Debido a que Ecuador es un área con características topográficas y climáticas muy variadas se eligió las tres ciudades más importantes Quito, Guayaquil y Cuenca para representarlo, adaptándonos a la información existente.

Los modelos de Redes Neuronales Recurrentes y Random Forest fueron implementados para predecir la concentración de los contaminantes NO_2 y SO_2 desde uno hasta cinco días de antelación para cada ciudad. A continuación, se presenta los siguientes resultados:

6.1. Análisis inicial para concentraciones de Dióxido de Azufre SO_2

Las siguientes figuras muestran la serie temporal de la variable respuesta y gráficos de dispersión por pares de variables, considerando la variable independiente (SO_2) y las dependientes de cada área de estudio.

6.1.1. Cuenca

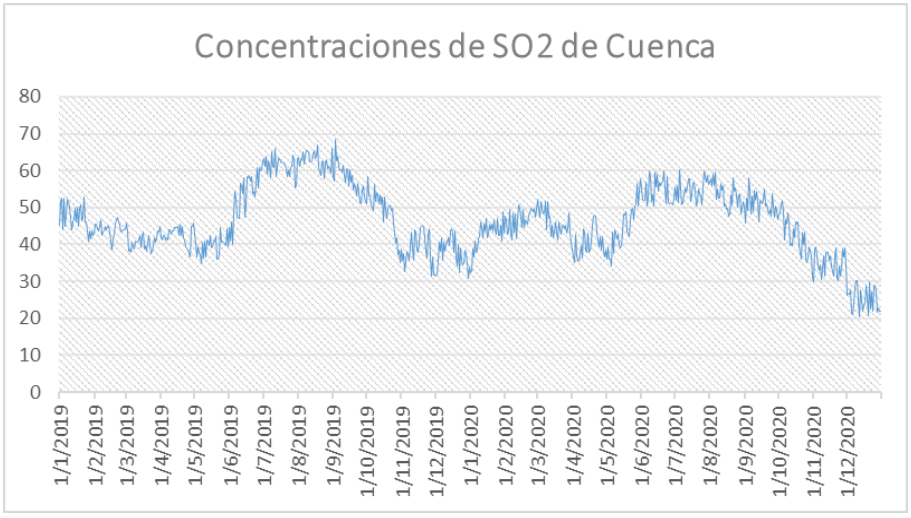

Gráfico 11. Concentraciones diarias de SO_2 de Cuenca.

En Cuenca se puede observar que existe una tendencia de aumento de concentraciones de dióxido de azufre en los meses de Julio y agosto y las concentraciones más bajas en los meses de noviembre hasta mayo.

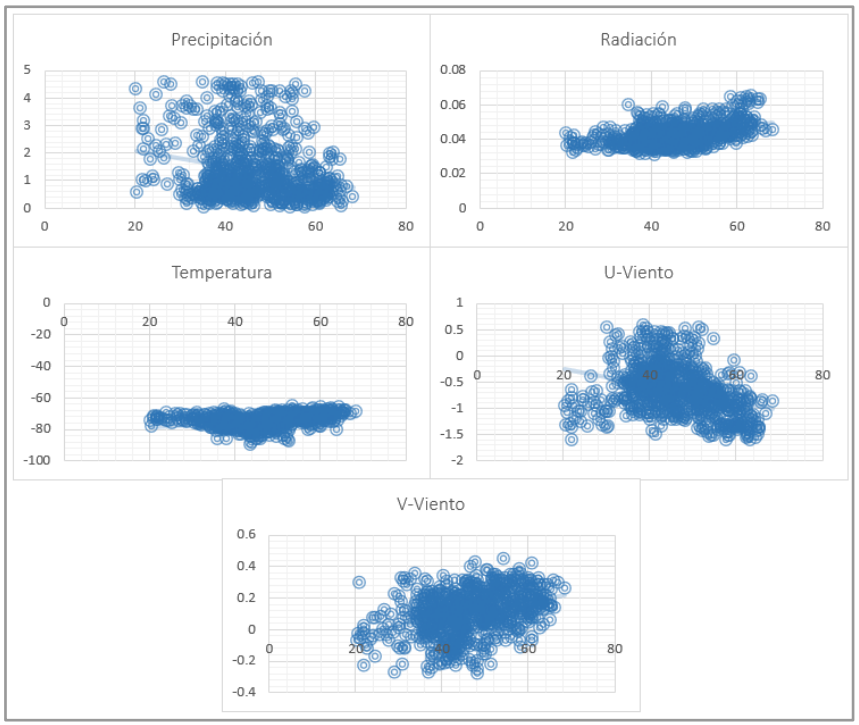

Gráfico 12. Dispersión de SO_2 de Cuenca con respecto a las variables meteorológicas.

La ilustración número cuatro nos demuestra que, para la gran mayoría de las variables utilizadas en el método, la distribución obtenida para la precipitación en relación con la concentración de SO_2 es la que más dispersa se encuentra. Y además observamos que la variable que posee una menor dispersión es la variable de temperatura puesto que los datos son más cercanos entre sí.

Tabla 4. Coeficiente de correlación lineal de la ciudad de Cuenca.

Coeficiente de correlación	
	SO2
Precipitación	*-0,230647533*
Radiación	*0,464741366*
Temperatura	*0,322406611*
U-Viento	*-0,325305037*
V-Viento	*0,3528424*

En la tabla a continuación se puede observar la correlación que tienen las variables meteorológicas con la concentración de la variable SO_2 en la ciudad de Cuenca. Podemos observar que la variable radiación es la que posee una correlación mayor positiva en comparación con las demás variables.

6.1.2. Guayaquil

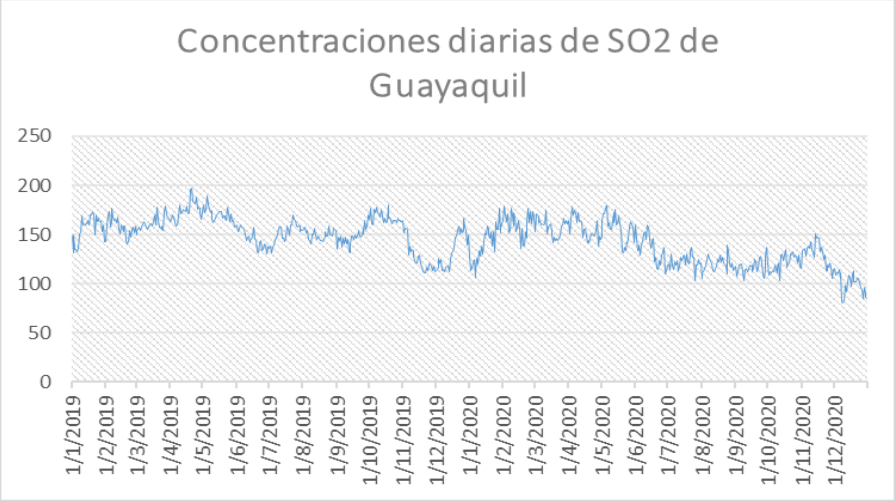

Gráfico 13. Concentraciones diarias de SO_2 de Guayaquil.

En la ciudad de Guayaquil se puede observar concentraciones más elevadas en comparación con la ciudad de Cuenca mostrando concentraciones más elevadas en los meses de marzo y abril en comparación con los meses de julio y agosto donde las concentraciones son menores donde las temperaturas son más bajas.

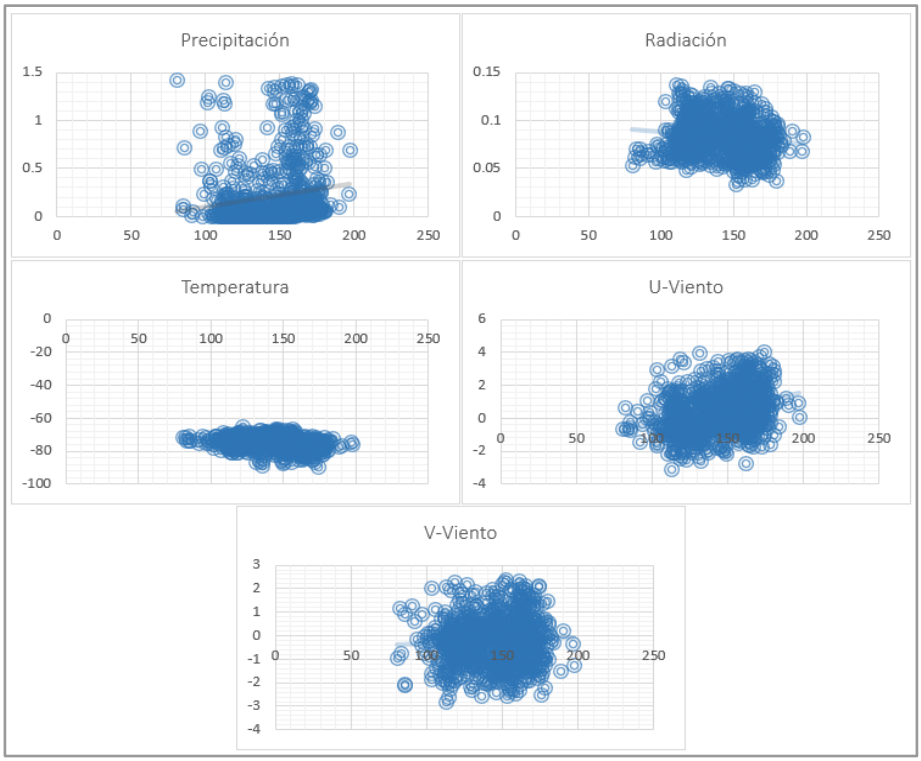

Gráfico 14. Dispersión de SO_2 de Guayaquil con respecto a las variables meteorológicas.

El comportamiento de las variables en Guayaquil es muy parecido a lo observado en Cuenca, con la diferencia que las variables de precipitación con la variable de concentración SO_2 se encuentran dispersas por lo tanto se puede considerar que existe una escasa relación entre ambas.

Tabla 5. Coeficiente de correlación lineal de la ciudad de Guayaquil.

Coeficiente de correlación	
	SO₂
Precipitación	*0,163442646*
Radiación	*-0,139379599*
Temperatura	*-0,318572022*
U-Viento	*0,324640863*
V-Viento	*0,055833386*

La siguiente tabla expresa los valores de correlación los cuales corresponden a la ciudad de Guayaquil y como se puede observar el mejor valor es el que se calcula con la componente u del viento lo que quiere decir que el viento que se mueve de Este a Oeste es el predominante.

6.1.3. Quito

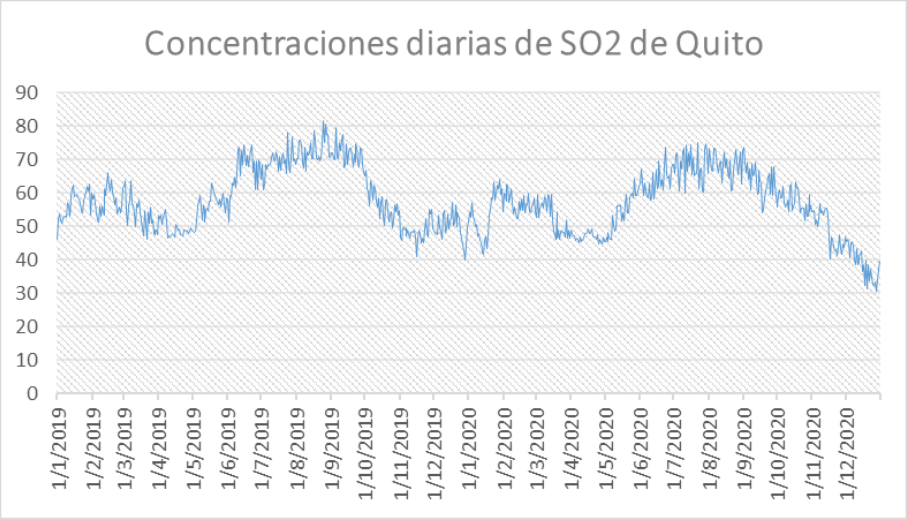

Gráfico 15. Concentraciones diarias de SO_2 de Quito.

De la misma manera en la ciudad de Quito se puede observar una tendencia a aumentar las concentraciones en temporadas de verano que consiste en junio, julio y agosto y las

concentraciones mínimas se presentan en los meses de diciembre hasta abril que corresponde a los meses de invierno.

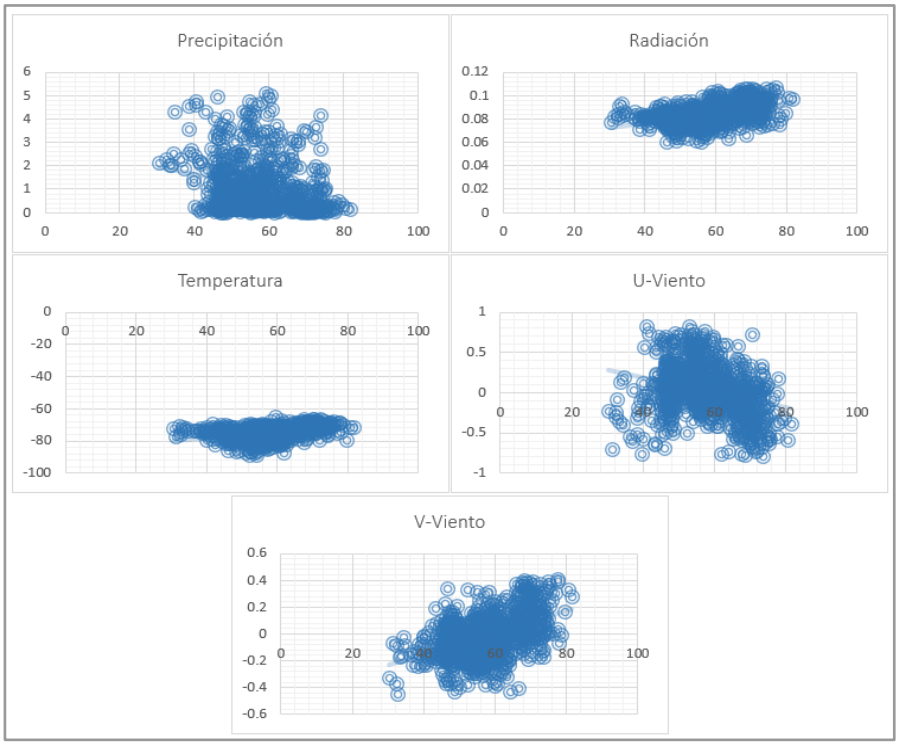

Gráfico 16. Dispersión de SO_2 de Quito con respecto a las variables meteorologías.

El comportamiento de las variables en Quito es muy parecido a lo observado en las áreas de estudio anteriores. Sin embargo, muestra la misma escasa relación con la variable de precipitación.

Tabla 6. Coeficiente de correlación lineal de la ciudad de Quito.

Coeficiente de correlación	
	SO2
Precipitación	*-0,262566488*
Radiación	*0,452826387*
Temperatura	*0,424300525*
U-Viento	*-0,284139627*
V-Viento	*0,46612052*

La siguiente tabla nos indica la correlación que poseen las distintas variables analizadas en comparación con las del SO_2 en la que el valor más alto son las que expresa la variable del componente V del viento, por lo tanto se puede expresar que el viento que se mueve de Norte a Sur es el predominante.

6.2. Análisis inicial para concentraciones de Dióxido de Nitrógeno NO_2

Las siguientes figuras muestran la serie temporal de la variable respuesta y gráficos de dispersión por pares de variables, considerando la variable independiente (NO_2) y las dependientes.

6.2.1. Cuenca

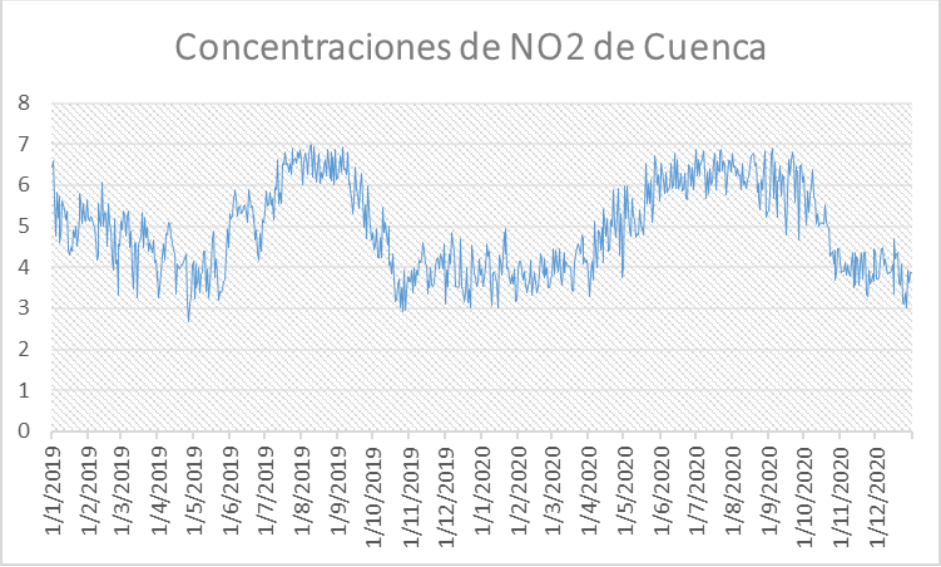

Gráfico 17. Concentraciones diarias de NO_2 de Cuenca.

Para el caso de los gases de NO_2 emitidos en la ciudad de Cuenca se puede observar una tendencia similar al de SO_2 donde muestra que las concentraciones van aumentando desde el mes de junio hasta septiembre y disminuyen desde los meses de octubre hasta abril.

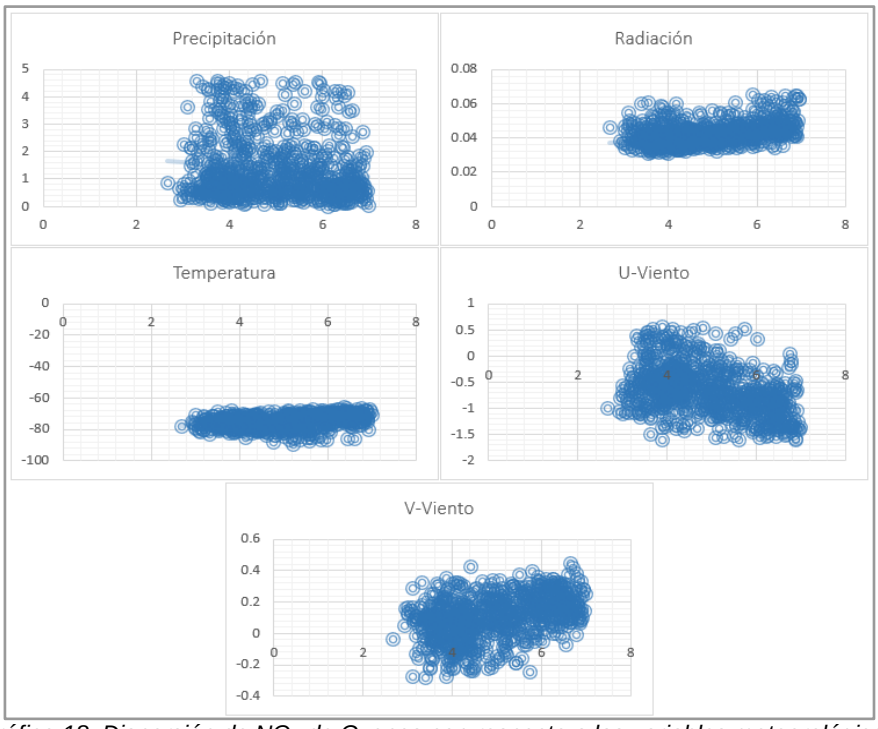

Gráfico 18. Dispersión de NO_2 de Cuenca con respecto a las variables meteorológicas.

Las distribuciones de las variables concuerdan con lo que se esperaba, es decir, cierta tendencia a describir distribuciones normales, y otras como la radiación con sesgo positivo y la precipitación con sesgo positivo aún más marcado. En los gráficos de la primera columna, se aprecia que las variables predictoras no presentan relación con la variable respuesta, los puntos están dispersos en toda la escala de las variables. Por otro lado, tampoco se observa relación entre las variables predictoras.

Tabla 7. Coeficiente de correlación lineal de la ciudad de Cuenca.

Coeficiente de Correlación	
	NO_2
Precipitación	-0,129592386
Radiación	0,40461519
Temperatura	0,349539158
U-Viento	-0,427519407
V-Viento	0,398884773

En la tabla 7 se puede apreciar la correlación entre cada variable meteorológicas con respecto a la variable de concentración de gases de NO_2, en la cual las variables de Radiación y Componente U del viento presentan una mejor relación mientras que la precipitación presenta una relación muy baja con respecto a la variable objetivo.

6.2.2. Guayaquil

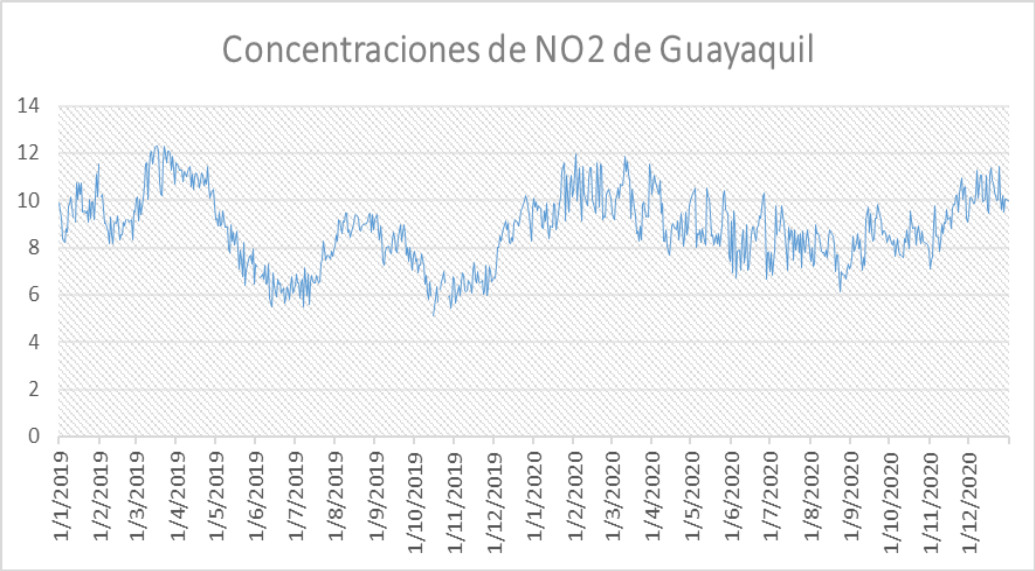

Gráfico 19. Concentraciones diarias de NO_2 de Guayaquil.

La tendencia de contaminación de gases de NO_2 en la ciudad de Guayaquil muestra un crecimiento de concentraciones en los meses de febrero, marzo y abril y concentraciones bajas en los meses de junio, julio y agosto.

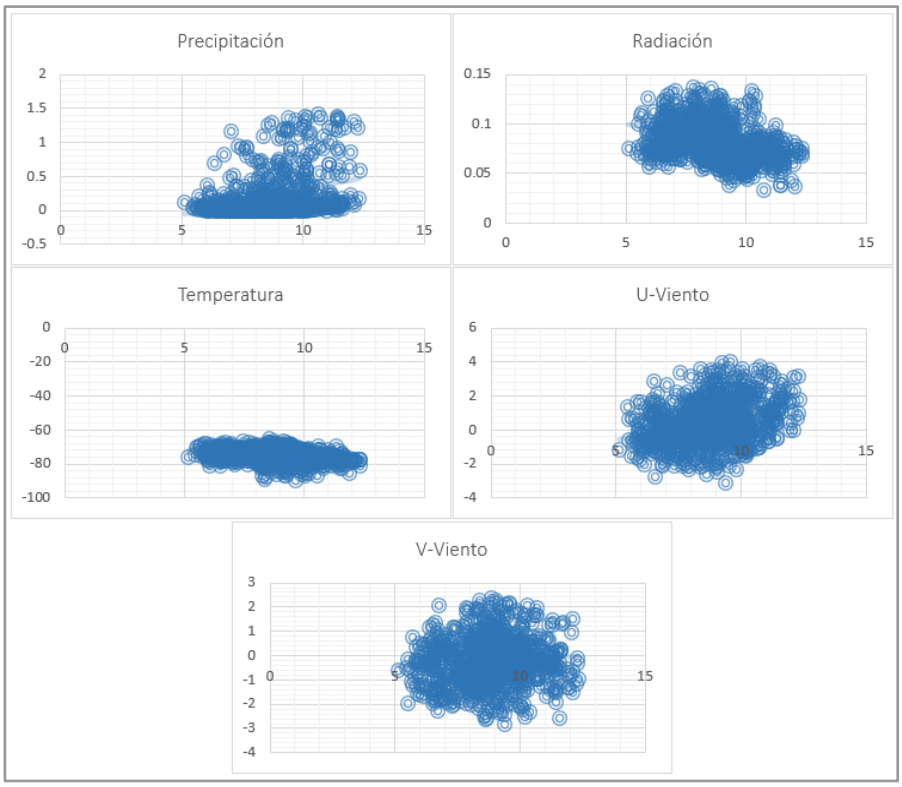

Gráfico 20. Gráfico de dispersión de NO_2 de Guayaquil con respecto a las variables meteorológicas.

El comportamiento de las variables en Guayaquil es muy parecido a lo observado en Cuenca.

Tabla 8. Coeficiente de correlación lineal de la ciudad de Guayaquil

Coeficiente de Correlación	
	NO_2
Precipitación	0,33887621
Radiación	-0,35449445
Temperatura	-0,31734489
U-Viento	0,31927359
V-Viento	0,07953216

Para el caso de Guayaquil, la relación entre variables es menor en comparación con la ciudad de Cuenca, mostrando una relación mayor con la radiación y precipitación, en lo que se refiere al viento podemos decir que el componente U tiene mayor relación con respecto al componente V lo que quiere decir que el viento que se mueve de Este a Oeste es el predominante.

6.2.3. Quito

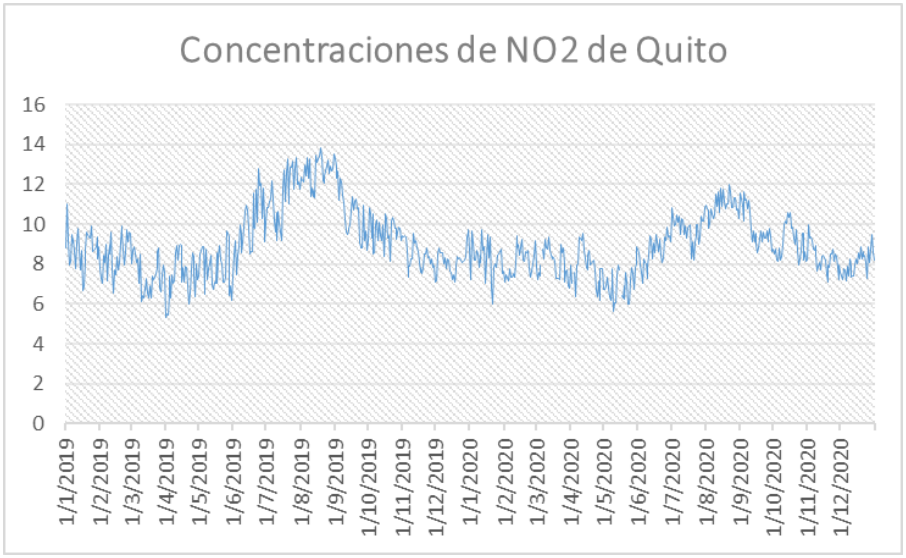

Gráfico 21. Concentraciones diarias de NO_2 de Quito.

El comportamiento de las variables en Quito es muy parecido a lo observado en Cuenca y Guayaquil.

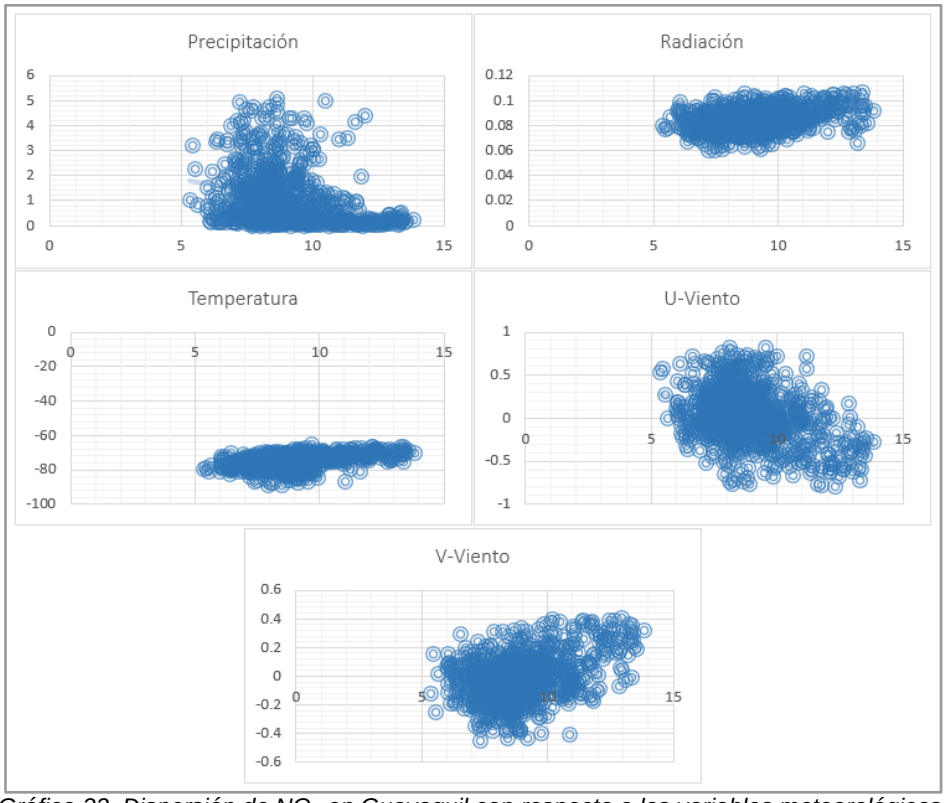

Gráfico 22. Dispersión de NO_2 en Guayaquil con respecto a las variables meteorológicas.

Con respecto a la ciudad de Quito, la relación entre las variables es baja, teniendo una mayor relación con las variables de temperatura radiación y Componente V del viento y como se pudo apreciar en las áreas de estudio anteriores la variable de precipitación es la que menor relación muestra con la variable de concentración.

Tabla 9. Coeficiente de Correlación lineal de la ciudad de Quito.

Coeficiente de Correlación	
	NO_2
Precipitación	-0,2675582
Radiación	0,36549768
Temperatura	0,48002824
U-Viento	-0,35380844
V-Viento	0,39445136

6.3. Análisis inicial de Variables Meteorológicas para NO_2 y SO_2

Las variables meteorológicas que se tomaron en cuenta para el pronóstico de emisiones de SO_2 y NO_2 fueron: Precipitación, Radiación, Temperatura y Componentes U y V del viento. En el Ecuador, estas variables cambian dependiendo la ciudad o la región en la que nos encontramos, por ello analizamos todas las variables de cada área de estudio.

6.3.1. Variables Meteorológicas en Cuenca

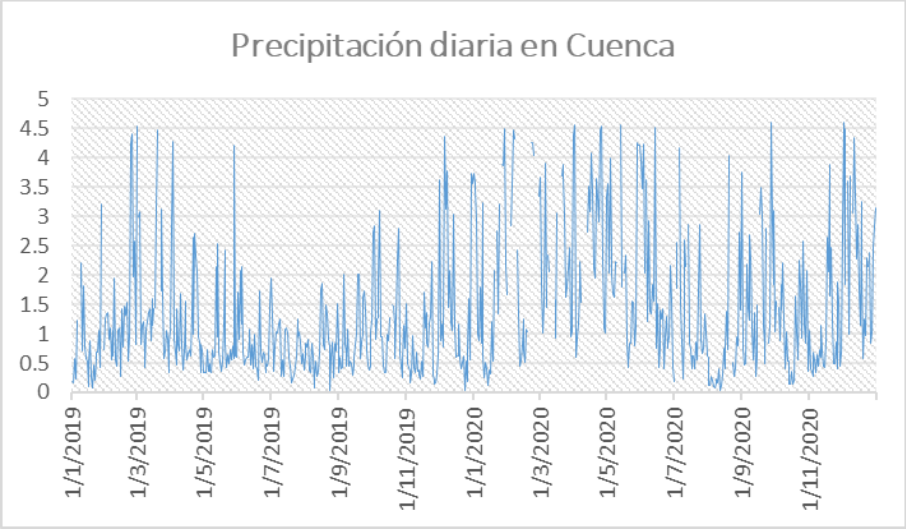

Gráfico 23. Precipitaciones diarias de Cuenca.

La figura anterior muestra la cantidad de lluvia que cae durante un día en la ciudad de Cuenca. puedes observar que la concentración de lluvia en los meses de enero hasta mayo es elevada con un aumento en la precipitación en el año 2020, con concentraciones máximas de precipitación de 4,59 mm en el mes de marzo y concentraciones mínimas de 0,025 mm en el mes de agosto.

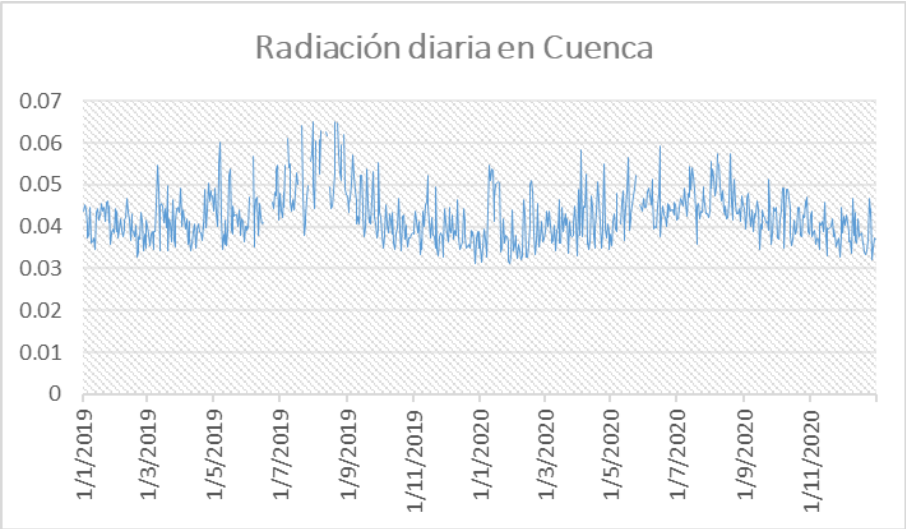

Gráfico 24. Radiación diaria de Cuenca.

La radiación es la energía emitida por el sol mediante ondas electromagnéticas, las mismas que muestran valores elevados para los meses secos de julio, agosto y septiembre, en comparación de los meses lluviosos de enero, febrero, marzo, abril y mayo. En el año 2019 la radiación fue más alta mostrando valores máximos de 0,065 Wm-2 al comparar con el año 2020 donde muestran valores máximos de radiación de 0,059 Wm-2 el mismo que presenta valores más bajos.

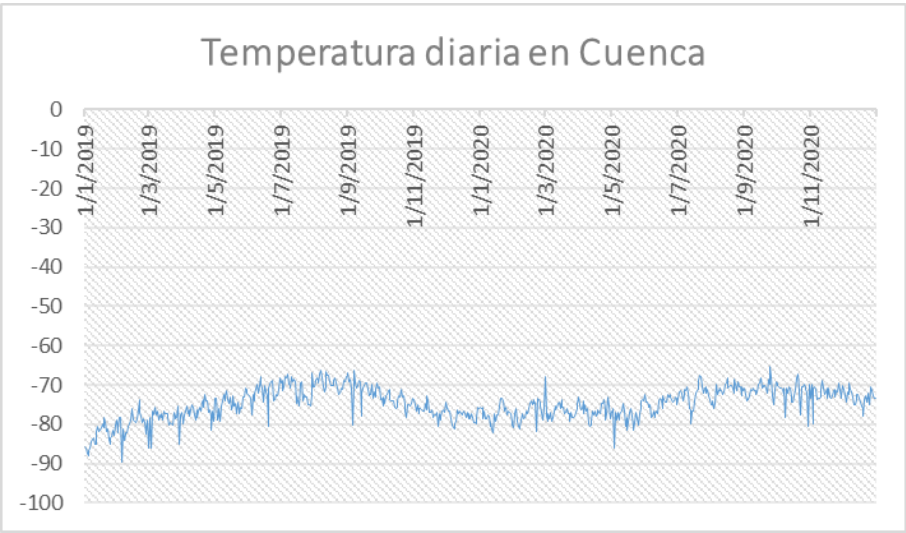

Gráfico 25. Temperatura diaria de Cuenca.

En el gráfico 25, los valores de temperatura diaria en la ciudad de Cuenca muestra valores de temperatura más altas en los meses desde junio hasta octubre y temperaturas más bajas en los meses de enero hasta marzo y noviembre y diciembre.

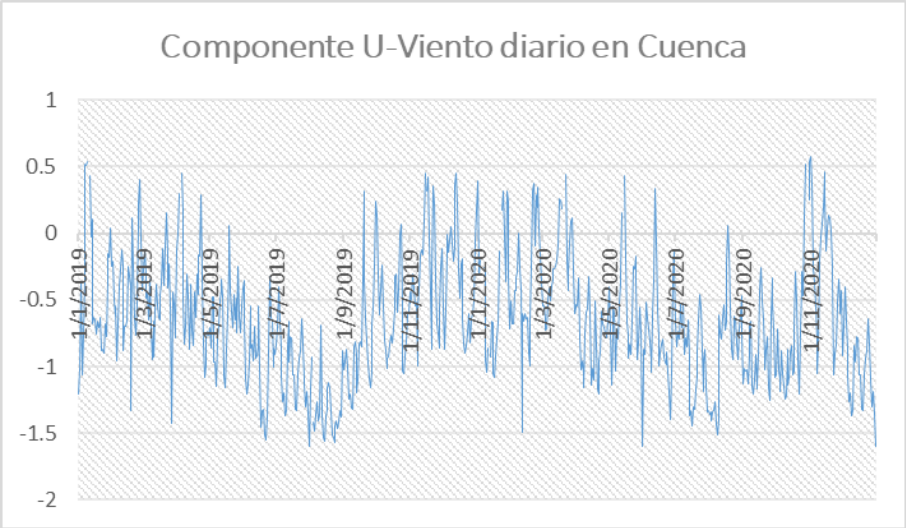

Gráfico 26. Componente U del viento diario de Cuenca.

El componente U del viento indica la velocidad del viento que se mueve desde el Este hacia el Oeste. En la figura muestra valores dispersos con valores mínimos en los meses de julio, agosto y septiembre lo que indica que el viento hacia el Este es predominante y valores más elevados en los demás meses del año, esto indica que el viento hacia el Oeste es el predominante.

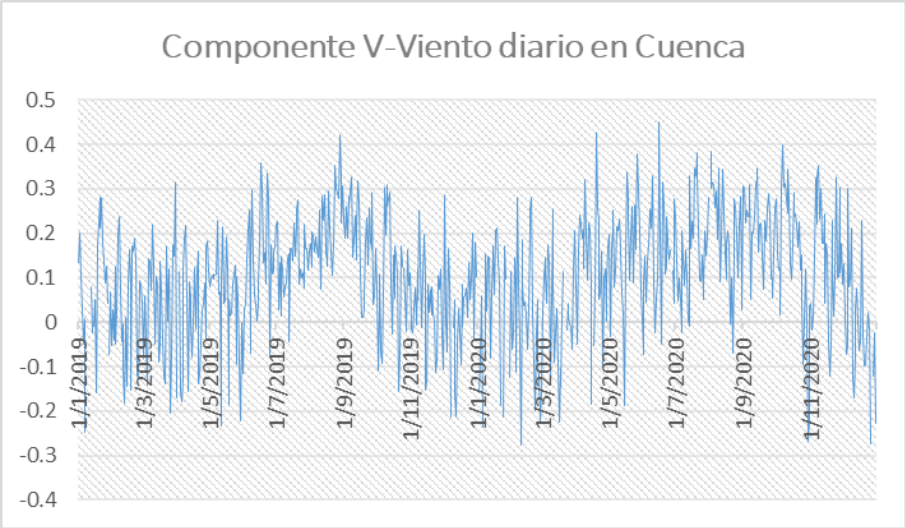

Gráfico 27. Componente V del viento diario de Cuenca.

El componente V del Viento es la velocidad del viento que se dirige desde el Norte hacia el Sur. En la figura muestra valores elevados con 0,44 m/s en los meses de julio, agosto y septiembre en los dos años de estudio a diferencia de los meses de enero hasta mayo y noviembre hasta diciembre con valores de -0,28 m/s.

6.3.2. Variables Meteorológicas en Guayaquil

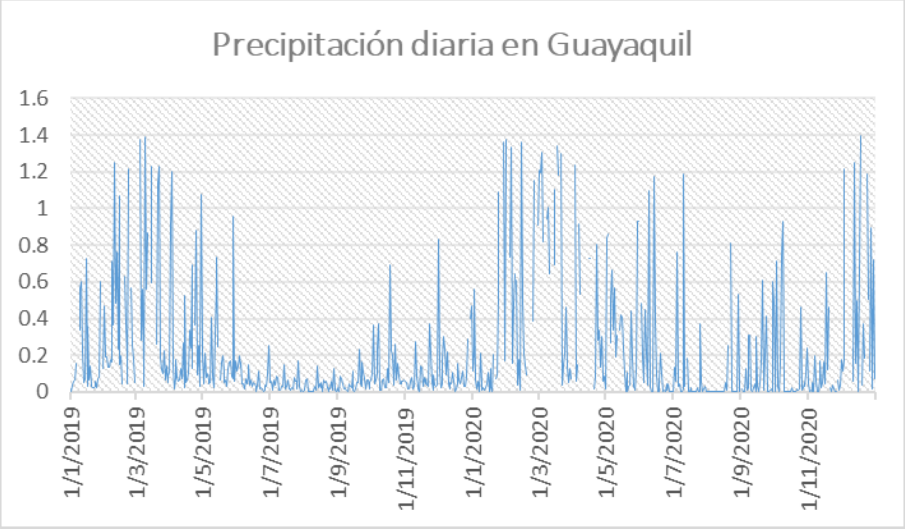

Gráfico 28. Precipitaciones diarias de Guayaquil.

La precipitación en la ciudad de Guayaquil es distinta a diferencia de las otras áreas de estudio por la ubicación de este punto. Aquí podemos apreciar que los meses con valores más elevados son los meses de enero hasta abril y los últimos dos meses noviembre y diciembre en el año 2019 y para el año 2020 la precipitación aumentó en el primer trimestre y el último trimestre.

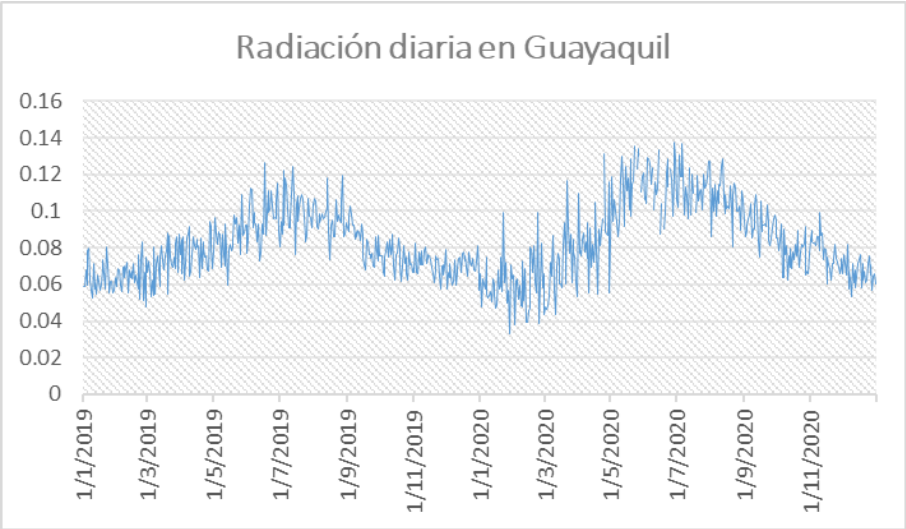

Gráfico 29. Radiación diaria de Guayaquil.

La radiación en Guayaquil muestra valores elevados con un máx.=0,13 Wm-2 lo que indica que la radiación es más alta en esta área de estudio. El valor mínimo e irradiación fue de 0,033 Wm-2 mostrando una tendencia marcada en los valores elevados en los meses de junio hasta octubre y los valores más pequeños en los meses de enero hasta marzo y septiembre hasta diciembre.

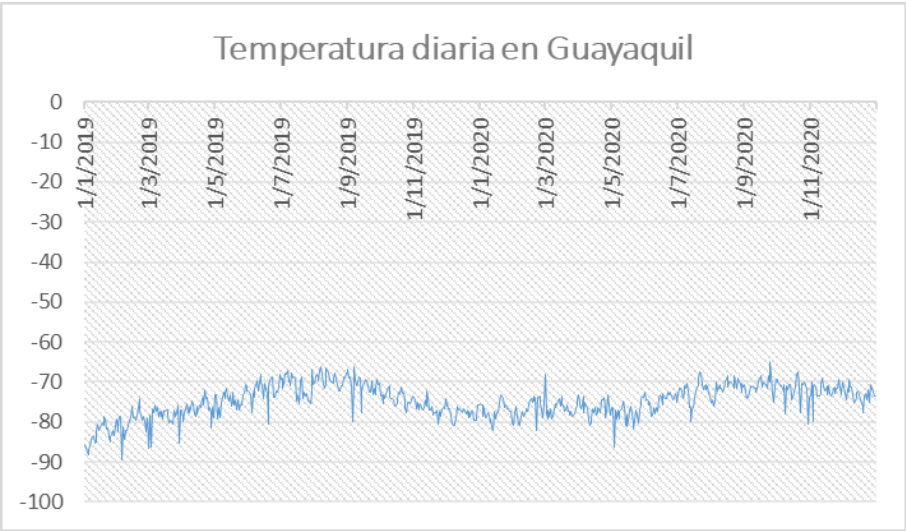

Gráfico 30. Temperatura diaria de Guayaquil.

La temperatura en el gráfico 30 muestra una tendencia más lineal ya que los valores no varían mucho a lo largo de todo el tiempo de estudio. En los meses de junio a agosto muestra cierta elevación en estos valores.

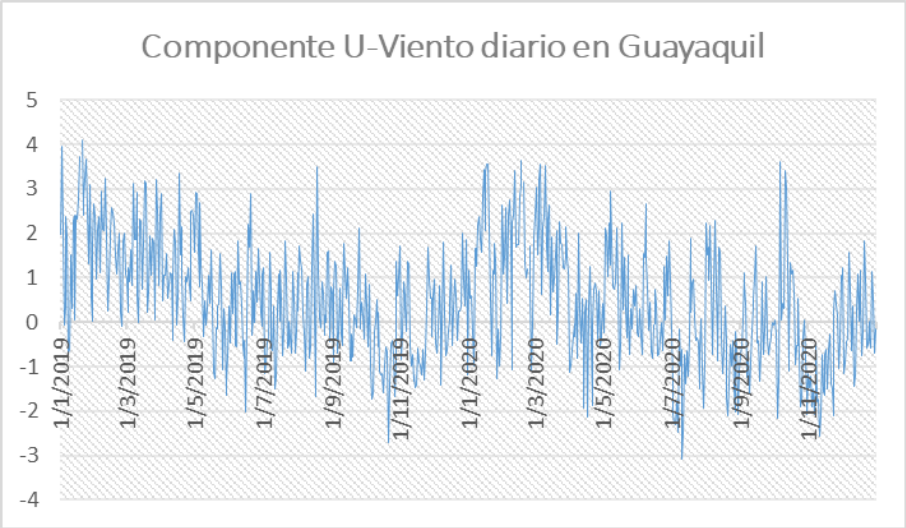

Gráfico 31. Componente U del viento diario de Guayaquil.

El gráfico 31 muestra la distribución temporal del componente U del Viento donde los valores más altos se registran en los meses de enero, febrero, marzo, abril y mayo con un máx. = 4,09 m/s hacia el este y un máx. = -3,08 m/s hacia el oeste. Cabe recalcar que el signo indicar la dirección del viento siendo así el signo positivo indica la dirección del viento hacia el este y el signo negativo indica que el viento se dirige hacia el oeste.

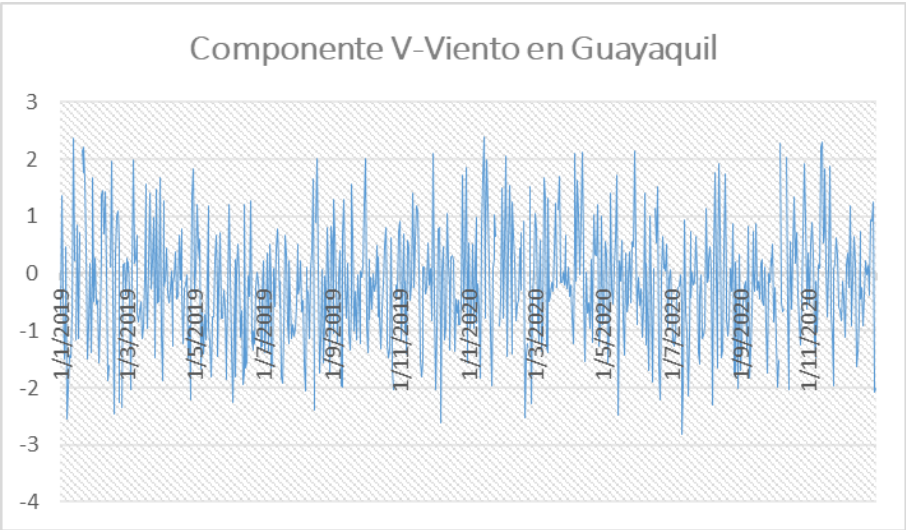

Gráfico 32. Componente V del viento diario de Guayaquil.

Con respecto al Componente V del Viento observamos en el gráfico anterior que no muestra una tendencia marcada pero muestran valores máximos de 2,38 m/s de Norte a Sur, además los valores máximos en sentido contrario de -2,81 m/s donde el signo negativo indica que los vientos van de Sur a Norte.

6.3.3. Variables Meteorológicas en Quito

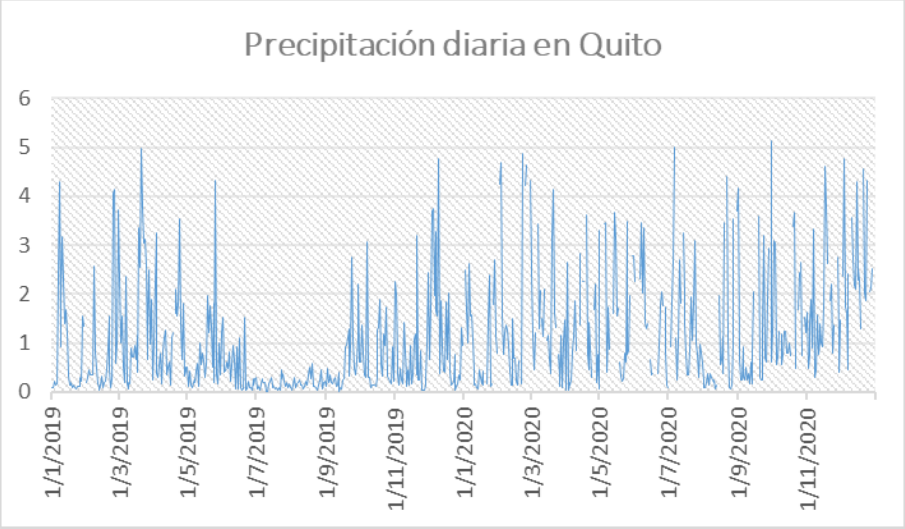

Gráfico 33. Precipitación diaria de Quito.

En la ciudad de Quito, la precipitación mostrada en el gráfico 33 se puede apreciar que los meses de enero hasta mayo los valores de precipitación son elevados con valores de 4,97 mm. Los meses donde menos lluvia presentan son los meses desde junio hasta septiembre con valores promedio de 0,34 mm, registrando los meses más secos. Para el año 2020 los valores de precipitación son más elevados con un máximo de 5,12 mm, cabe recalcar que para este año los valores de precipitación se mantuvieron elevados.

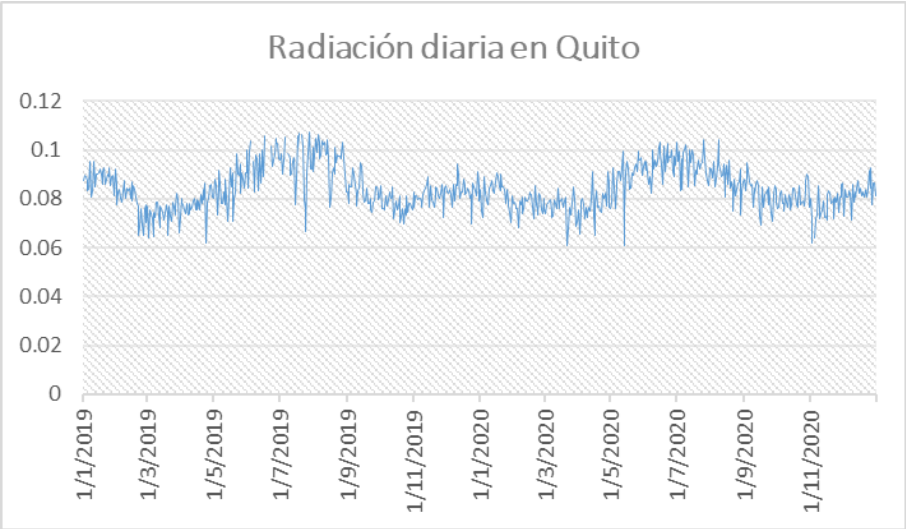

Gráfico 34. Radiación diaria de Quito.

El gráfico 34 mostró los valores de radiación a través de todo el tiempo de estudio donde muestra una tendencia a mostrar valores más elevados de máx.=0,11 Wm-2 que corresponden a los meses más secos desde junio hasta septiembre. Mientras que los meses que muestran valores más bajos son octubre, noviembre y diciembre con valores de 0,061 Wm-2.

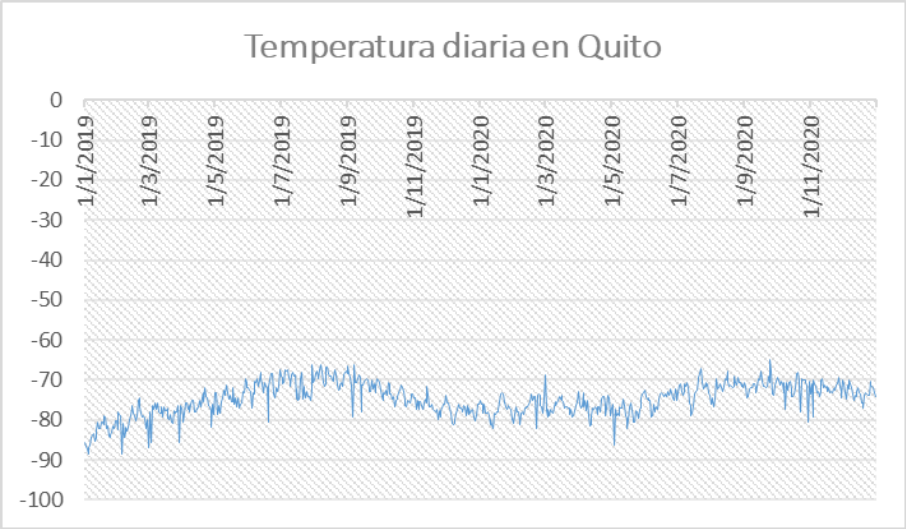

Gráfico 35. Temperatura diaria de Quito.

Los valores en el gráfico 35 corresponden a la temperatura registrada en Quito durante los años 2019 y 2020. En la misma podemos observar que los valores más altos de temperatura se pueden apreciar en los meses de junio, julio, agosto y septiembre.

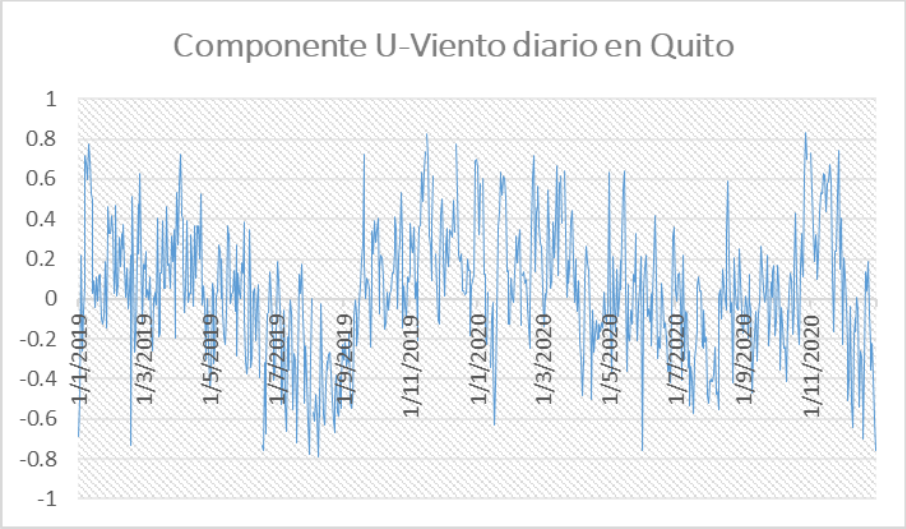

Gráfico 36. Componente U del viento diario de Quito.

En lo que respecta al Componente U del viento en Quito podemos observar en el gráfico 36 que los valores más elevados son de 0,83 m/s desde el Este hacia el Oeste y -0,79 m/s que va en sentido contrario. Esto quiere decir que los meses de enero hasta mayo y octubre hasta diciembre los vientos se dirigen hacia el oeste, a diferencia de los meses de junio, julio, agosto y septiembre los vientos se dirigen hacia el Este.

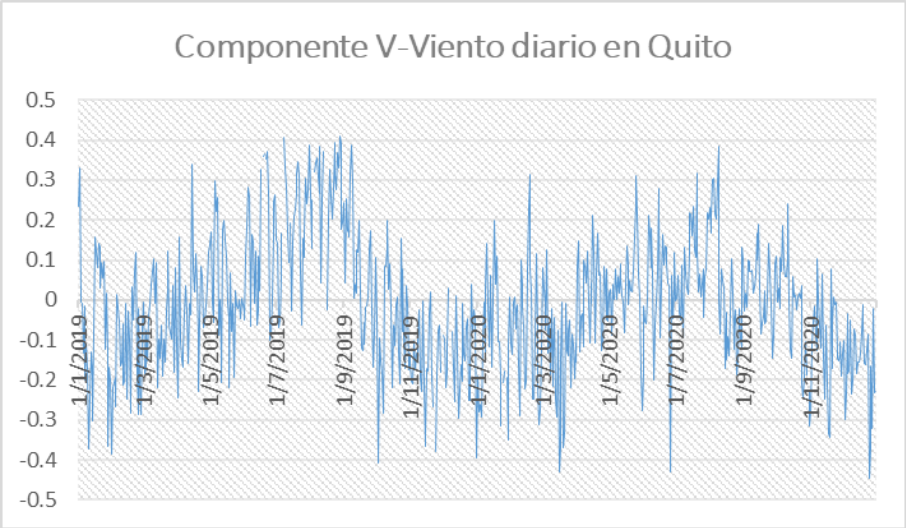

Gráfico 37. Componente V del viento diario de Quito.

El gráfico anterior muestra la dirección y velocidad del viento de Norte a Sur y viceversa, donde, los valores más elevados 0,41 m/s en sentido de Norte a Sur corresponden a los meses de junio, julio, agosto y septiembre y los más altos 0,44 m/s en sentido de Sur a Norte que se registran desde octubre hasta marzo.

6.4. Resultados sobre la variable SO_2 y NO_2

6.4.1. Redes neuronales

Se utilizó la normalización Zscore en las variables independientes para que faciliten el entrenamiento de las redes. Se utilizó el optimizador Adam con el hiperparámetro learning_rate=0,001. En general el entrenamiento de las redes neuronales se realizó con 1000 épocas, es decir 1000 pasadas del conjunto de datos completo para ajustar los pesos de la red. Todo fue realizado en el entorno Google Colab que usa el lenguaje Python, y las librerías Keras y Tensorflow para facilitar la construcción de las redes.

Para las 3 ciudades se entrenaron varias configuraciones de redes neuronales siguiendo el siguiente diseño experimental.

Tabla 10. Tabla de diseño experimental de redes neuronales.

Capas ocultas	*1*
Función de activación	ReLU
Número de repeticiones de cada configuración	30
Número de épocas	1000
Número de neuronas en la capa oculta	16, 32, 64
Número de retrasos	0, 1, 2, 3, 4
Días hacia adelante para la predicción	1, 2, 3, 4, 5

Las siguientes figuras muestran el progreso del entrenamiento de la red en función de las épocas, para la predicción de SO_2 para 1 día hacia adelante. La arquitectura de la red utilizada fue: 1 capa oculta, 64 neuronas, y un número de retrasos de 2.

6.4.1.1. Predicción del SO_2 mediante Redes Neuronales Recurrentes.

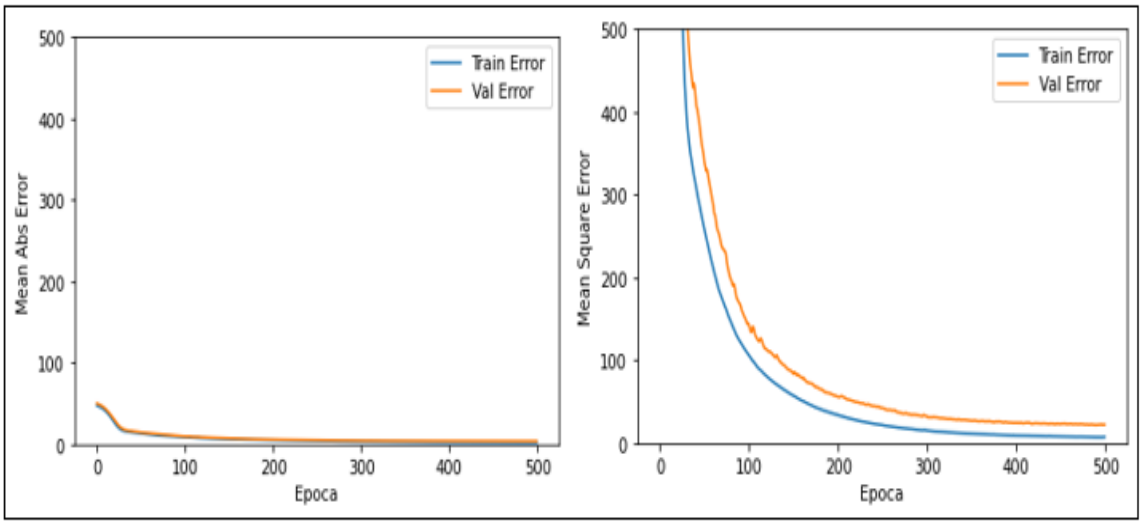

Gráfico 38. Entrenamiento de la red neuronal en función de las épocas del SO_2.

Se puede apreciar que conforme avanza el entrenamiento, el error en el conjunto de entrenamiento sigue decreciendo. Sin embargo, alrededor de la época 300 el error en el conjunto de validación empieza a aumentar ligeramente, lo cual es un indicador de que está iniciando la etapa de sobreajuste (overfitting), por lo que se debería detener el entrenamiento alrededor de tal época. Se usó la función *EarlyStopping* de Keras para detener el entrenamiento cuando el error de validación empezó a subir. Las siguientes figuras muestran el avance y finalización del entrenamiento.

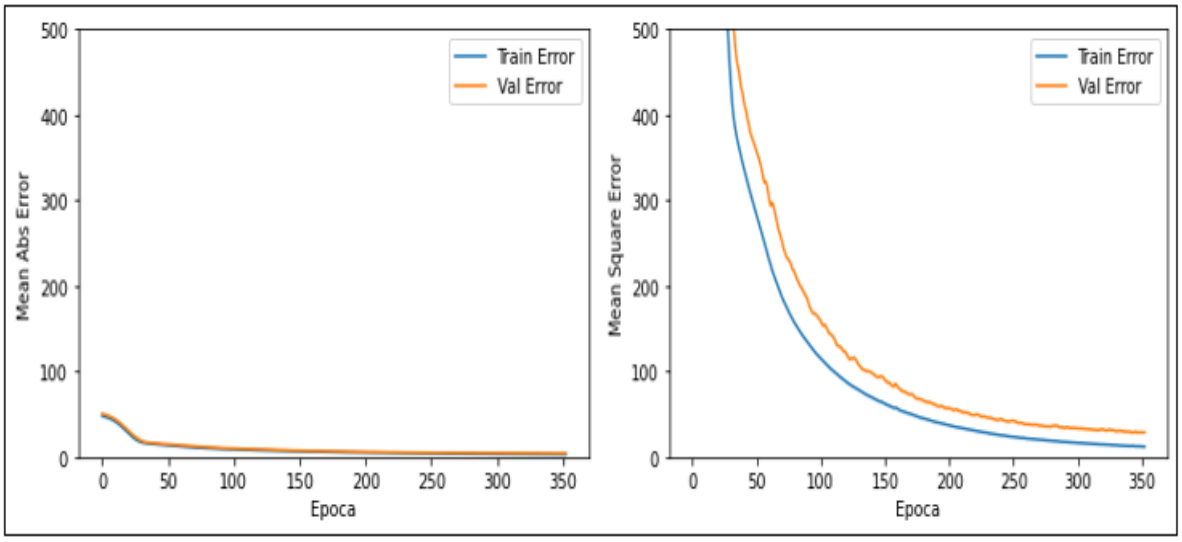

Gráfico 39. Entrenamiento de la red neuronal aplicando detección temprana del error de validación del SO_2.

El gráfico 39 muestran que el entrenamiento se detuvo en la época 350 y que en este punto se obtuvo la red con el entrenamiento óptimo.

A continuación, en el gráfico 40 muestra la relación entre los valores de SO_2 verdaderos y los valores predichos por la red entrenada para el conjunto de prueba.

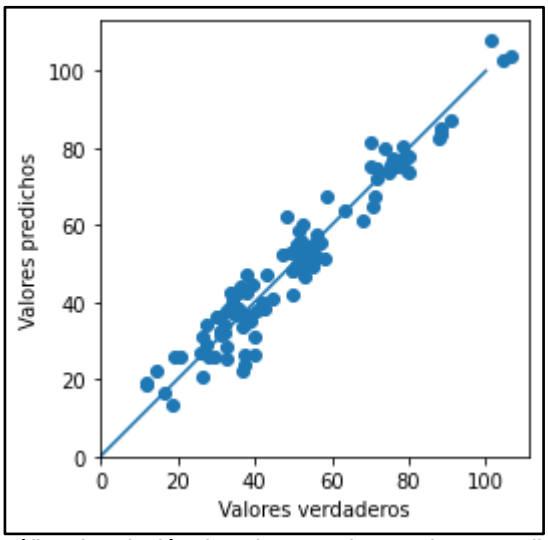

Gráfico 40. Gráfico de relación de valores reales y valores predichos del SO_2.

En un ajuste perfecto, los puntos caerían sobre la línea azul. Analizamos los residuos entre el valor verdadero y el valor predicho por la red. Se puede observar que existe una relación entre los valores reales y los valores predichos.

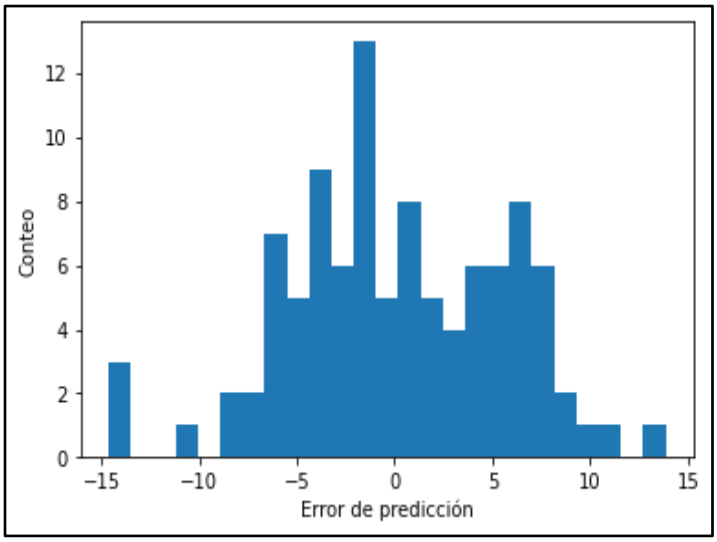

Gráfico 41. Histograma del error de predicción del SO_2.

Los valores de los errores tienden a seguir una distribución normal y el valor central tiende a 0. Este es un indicador de que el modelo está capturando algún patrón entre las variables independiente y la dependiente.

El proceso anterior se repitió 30 veces con una inicialización diferente de los pesos de la red y con una partición diferente del conjunto inicial para el entrenamiento y la prueba. Las siguientes tablas muestran los resultados del promedio y desviación estándar de las métricas de evaluación de bondad de ajuste (R^2 o coeficiente de determinación) y de error (MAPE, RMSE y PBIAS).

Tabla 11. Resultados de análisis de Redes Neuronales Recurrentes para SO_2 de Cuenca.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.08	0.01	0.08	0.01	4.06	0.28	0.69	0.5	0.96	0	0.08	0.01	4.06	0.27	0.67	0.56	**0.96**	**0**	**0.08**	**0.01**	**4.06**	**0.23**	**0.7**	**0.56**
		1	0.96	0.01	0.1	0.01	4.82	0.31	1.24	0.84	0.96	0.01	0.1	0.01	4.76	0.35	1.09	0.73	0.96	0.01	0.1	0.01	4.76	0.34	1.14	0.9
		2	0.95	0.03	0.1	0.02	5.23	1.03	1.22	0.88	0.95	0.02	0.1	0.02	5.22	0.9	1.06	0.83	0.95	0.03	0.1	0.02	5.29	1.07	1.09	0.83
		3	0.83	0.05	0.1	0.02	5.43	0.89	1.2	0.85	0.83	0.09	0.1	0.02	5.63	0.98	1.04	1.07	0.88	0.04	0.1	0.01	5.28	0.55	0.95	0.6
		4	0.74	0.09	0.12	0.01	6.69	0.85	1.69	1.45	0.8	0.06	0.12	0.01	6.32	0.73	1.5	1.1	0.85	0.03	0.11	0.01	6.02	0.56	1.15	0.89
	2	0	0.95	0	0.08	0.01	4.59	0.28	0.7	0.57	0.95	0	0.08	0.01	4.6	0.3	0.72	0.55	**0.95**	**0**	**0.08**	**0.01**	**0.08**	**0.01**	**0.08**	**0.01**
		1	0.73	0.57	0.94	0.01	0.11	0.01	5.62	0.41	1.36	1.06	0.94	0.01	0.11	0.01	5.72	0.55	1.38	1.15	0.94	0.01	0.11	0.01	5.62	0.43
		2	1.35	0.97	0.82	0.05	0.09	0.01	5	0.66	1.14	1.05	0.84	0.05	0.09	0.01	4.98	0.7	1.17	0.89	0.85	0.03	0.09	0.01	4.87	0.58
		3	0.81	0.11	0.11	0.02	5.76	1.08	1.15	0.81	0.85	0.05	0.1	0.01	5.51	0.82	1.22	0.92	0.87	0.04	0.1	0.01	5.49	0.64	1.14	0.85
		4	0.81	0.11	0.13	0.02	6.8	0.98	1.58	1.3	0.86	0.06	0.12	0.02	6.61	0.97	1.89	1.24	0.87	0.05	0.12	0.02	6.71	0.83	1.62	1
	3	0	0.94	0	0.1	0.01	5.29	0.35	0.78	0.69	0.94	0	0.1	0.01	5.28	0.38	0.78	0.68	**0.94**	**0**	**0.1**	**0.01**	**5.23**	**0.37**	**0.8**	**0.65**
		1	0.92	0.01	0.13	0.01	6.6	0.43	1.34	1.32	0.92	0.01	0.13	0.02	6.45	0.45	1.29	1.25	0.92	0.01	0.13	0.02	6.53	0.49	1.37	1.29
		2	0.76	0.11	0.1	0.02	5.56	0.85	1.05	1.02	0.81	0.07	0.1	0.01	5.32	0.67	0.96	0.96	0.84	0.05	0.1	0.01	5.08	0.59	1.03	0.89
		3	0.76	0.14	0.12	0.02	6.24	1.15	1.55	1.23	0.82	0.08	0.11	0.02	6.08	0.9	1.48	1.03	0.87	0.04	0.1	0.01	5.58	0.7	1.07	0.74
		4	0.79	0.11	0.13	0.02	6.91	0.93	1.95	1.58	0.8	0.14	0.13	0.02	7.09	1.27	1.68	1.58	0.85	0.05	0.12	0.01	6.7	0.71	1.53	1.2
	4	0	0.92	0.01	0.11	0.01	5.84	0.42	0.69	0.66	0.92	0.01	0.11	0.01	5.79	0.45	0.82	0.74	**0.92**	**0.01**	**0.11**	**0.01**	**5.84**	**0.44**	**0.74**	**0.74**
		1	0.9	0.01	0.14	0.02	7.04	0.48	1.79	1.07	0.9	0.01	0.14	0.02	7.18	0.43	1.64	1.16	0.91	0.01	0.14	0.02	7.02	0.52	1.61	1.18
		2	0.76	0.08	0.11	0.02	5.72	0.96	1.24	1.06	0.79	0.09	0.11	0.02	5.62	0.9	1.24	1.12	0.81	0.09	0.1	0.02	5.47	0.95	1.21	1.2
		3	0.76	0.07	0.12	0.02	6.42	0.81	1.64	1.13	0.81	0.06	0.12	0.02	6.34	0.87	1.41	0.95	0.84	0.04	0.11	0.01	6.03	0.69	1.25	0.9
		4	0.73	0.17	0.14	0.03	7.53	1.27	1.81	1.4	0.81	0.08	0.13	0.02	7.1	0.87	1.95	1.21	0.85	0.07	0.13	0.02	7.12	0.92	1.66	1.11
	5	0	0.91	0.01	0.11	0.01	6.16	0.48	0.58	0.51	0.91	0.01	0.11	0.01	6.19	0.44	0.71	0.49	**0.91**	**0.01**	**0.11**	**0.01**	**6.19**	**0.47**	**0.68**	**0.5**
		1	0.88	0.01	0.16	0.02	7.69	0.67	1.77	1.17	0.89	0.01	0.16	0.02	7.6	0.53	1.7	1.19	0.89	0.01	0.16	0.02	7.62	0.45	1.55	1.13
		2	0.74	0.11	0.11	0.02	5.86	0.85	1.21	0.81	0.76	0.08	0.11	0.02	5.84	0.88	1.25	1.09	0.79	0.09	0.11	0.02	5.64	0.81	1.28	0.9
		3	0.74	0.09	0.12	0.01	6.69	0.85	1.69	1.45	0.8	0.06	0.12	0.01	6.32	0.73	1.5	1.1	0.85	0.03	0.11	0.01	6.02	0.56	1.15	0.89
		4	0.78	0.08	0.14	0.02	7.37	1	1.67	1.23	0.81	0.09	0.14	0.02	7.29	0.97	1.85	1.25	0.83	0.12	0.13	0.02	7.07	1.05	1.41	1.08

En la tabla muestra el resultado de las diferentes arquitecturas donde se puede decir que el modelo que mejor resultados mostró fue con la arquitectura de 64 neuronas, ahora bien, teniendo en cuenta el número de retrasos observamos que mientras menor sea el número de retraso mejores son los valores de validación, teniendo el retraso óptimo para todos los modelos un valor de 0. Además la bondad de ajuste va disminuyendo de acorde los días de predicción vayan aumentando, se tomó en cuenta todas la métricas de validación para elegir el mejor resultados tendiendo como la mejor predicción para el primer día con un R^2=0,96, MAPE=0,08, RMSE=4,06 y PBIAS=0,7, seguido del día 2 con un R^2=0,95, MAPE=0,08, RMSE=0,08 y PBIAS=0.08 obteniendo el mejor porcentaje de subestimación menor, el día 3 muestra un R^2=0,94, MAPE=0,1, RMSE=5,23 y PBIAS=0,8 lo que indica que la bondad de ajuste es buena pero el RMSE aumenta, para el día 4 tenemos un R^2=0,92, MAPE=0,11, RMSE=5,81 y PBIAS=0,74 y para el día 5 tenemos un R^2=0,91, MAPE=0,11, RMSE=6,19 y PBIAS=0,68 donde se puede observar que el valor de RMSE es el más alto.

Tabla 12. Resultados de análisis de Redes Neuronales Recurrentes para SO_2 de Guayaquil.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.86	0.01	0.05	0	8.55	0.7	0.52	0.39	0.87	0.01	0.05	0	8.56	0.63	0.54	0.32	**0.87**	**0.01**	**0.05**	**0**	**8.49**	**0.57**	**0.55**	**0.4**
		1	0.87	0.01	0.05	0	9.34	0.74	0.65	0.46	0.88	0.01	0.05	0	9.26	0.65	0.61	0.49	0.88	0.01	0.05	0	9.28	0.69	0.74	0.5
		2	0.81	0.03	0.06	0.01	11.8	1.22	0.73	0.59	0.87	0.02	0.06	0.01	11.3	1.13	0.61	0.61	0.89	0.02	0.06	0.01	11.1	1.2	0.67	0.67
		3	0.77	0.05	0.08	0.01	14.9	1.7	1.09	0.77	0.85	0.03	0.08	0.01	13.6	1.49	0.87	0.72	0.87	0.04	0.08	0.01	13.5	1.28	0.86	0.72
		4	0.25	0.4	0.13	0.03	23.6	5.17	2.79	1.87	0.38	0.37	0.12	0.03	22.3	5.11	2.62	2.34	0.48	0.38	0.12	0.03	21.1	5.54	2.26	1.58
	2	0	0.81	0.01	0.06	0	10.2	0.65	0.58	0.45	0.82	0.01	0.06	0	10.2	0.58	0.56	0.4	0.82	0.01	0.06	0	10.1	0.52	0.57	0.39
		1	0.83	0.01	0.06	0	10.5	0.76	0.67	0.52	0.84	0.01	0.06	0	10.8	0.71	0.7	0.52	**0.85**	**0.02**	**0.06**	**0.01**	**10.7**	**0.78**	**0.69**	**0.62**
		2	0.8	0.03	0.07	0.01	12.7	1.19	0.7	0.57	0.84	0.02	0.07	0.01	12.3	0.97	0.83	0.64	0.86	0.02	0.07	0.01	12	0.96	0.78	0.57
		3	0.73	0.04	0.09	0.01	15.9	1.93	1.25	0.9	0.83	0.03	0.08	0.01	14.4	1.63	1.1	0.78	0.85	0.03	0.08	0.01	14.4	1.42	0.99	0.76
		4	0.2	0.41	0.13	0.03	24.5	5.74	3.55	2.26	0.22	0.42	0.13	0.03	24.4	5.7	3.16	2.4	0.38	0.34	0.12	0.02	22.9	4.8	2.76	2.06
	3	0	0.78	0.01	0.06	0	10.6	0.7	0.6	0.39	0.79	0.01	0.06	0	10.6	0.7	0.55	0.39	0.79	0.01	0.06	0	10.6	0.75	0.57	0.43
		1	0.79	0.02	0.06	0.01	11.6	1.05	0.8	0.74	0.81	0.02	0.06	0	11.5	0.8	0.84	0.67	**0.83**	**0.02**	**0.06**	**0**	**11.3**	**0.83**	**0.83**	**0.63**
		2	0.77	0.03	0.07	0.01	13.4	1.3	0.78	0.75	0.82	0.02	0.07	0.01	12.8	1.02	0.81	0.63	0.84	0.02	0.07	0	12.8	0.89	0.75	0.71
		3	0.72	0.05	0.09	0.01	16.7	1.71	1.42	1.17	0.81	0.04	0.08	0.01	15.1	1.62	1.29	0.86	0.84	0.04	0.08	0.01	14.7	1.36	1.12	0.71
		4	0.15	0.41	0.14	0.03	25.4	5.38	3.4	2.47	0.27	0.44	0.13	0.02	24.1	4.99	3.36	2.42	0.38	0.35	0.12	0.02	23.2	4.71	2.94	2.23
	4	0	0.74	0.01	0.06	0	11.6	0.63	0.7	0.56	0.75	0.01	0.06	0	11.5	0.58	0.66	0.54	0.75	0.01	0.06	0	11.5	0.62	0.68	0.52
		1	0.75	0.02	0.07	0	12.1	0.91	0.79	0.58	0.78	0.02	0.07	0	12	0.76	0.79	0.69	**0.8**	**0.02**	**0.06**	**0**	**11.9**	**0.64**	**0.83**	**0.59**
		2	0.74	0.03	0.08	0.01	13.9	1.54	0.82	0.53	0.8	0.02	0.07	0.01	13.2	1.2	0.7	0.8	0.82	0.03	0.07	0.01	13.2	1.19	0.78	0.87
		3	0.69	0.05	0.09	0.01	17.3	2.25	1.46	1.01	0.78	0.03	0.09	0.01	16	1.57	1.15	0.91	0.82	0.04	0.09	0.01	15.5	1.42	1.02	0.79
		4	-0	0.49	0.15	0.03	27	5.25	3.84	2.43	0.27	0.37	0.13	0.03	24.3	5.02	2.9	1.95	0.4	0.36	0.13	0.03	23.2	5.32	2.54	1.59
	5	0	0.71	0.01	0.07	0.01	12.3	0.85	0.71	0.52	0.72	0.01	0.07	0.01	12.4	1.07	0.72	0.61	0.72	0.01	0.07	0.01	12.4	0.96	0.67	0.57
		1	0.74	0.02	0.07	0.01	12.9	0.97	0.82	0.63	0.76	0.02	0.07	0.01	12.8	1.09	0.76	0.68	**0.79**	**0.02**	**0.07**	**0.01**	**12.6**	**0.97**	**0.77**	**0.71**
		2	0.71	0.04	0.08	0.01	14.9	1.2	0.89	0.61	0.77	0.03	0.08	0.01	14.6	1.13	0.86	0.69	0.8	0.03	0.08	0.01	14	1.12	0.79	0.58
		3	0.64	0.09	0.1	0.01	18.3	1.88	1.63	1.07	0.73	0.06	0.1	0.01	17.1	1.79	1.1	0.92	0.8	0.05	0.09	0.01	16.4	1.66	1.02	0.71
		4	-0.1	0.46	0.15	0.03	28.4	4.68	4.17	2.62	0.06	0.35	0.14	0.02	26.5	4.23	3.38	2.26	0.33	0.34	0.13	0.02	24.1	4.56	2.64	1.87

Según el modelo de predicción, para Guayaquil presentan una bondad de ajuste más baja que en Cuenca. Podemos decir que el número de neuronas influye de manera importante ya que mientras menos neuronas tenga la bondad de ajuste disminuye al igual que el número de retrasos debido a que mientras más números de retrasos presenta el modelo menor es la bondad de ajuste. Con respecto a la mejor predicción hecha se puede considera que el día 1 fue la mejor predicción con un R^2=0,87, MAPE=0,05, RMSE=8,49 y PBIAS=0,4, seguido de un R^2=0,85, MAPE=0,06, RMSE=10,7 y PBIAS=0,69 y se mantiene disminuyendo para los días siguientes pronosticados. Los valores más bajos registrados corresponden a los retrasos de valor 4.

Tabla 13. Resultados de análisis de Redes Neuronales Recurrentes para SO_2 de Quito.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.84	0.01	0.06	0	4.04	0.26	0.43	0.37	0.85	0.01	0.06	0	4.03	0.27	0.46	0.38	0.85	0.01	0.06	0	4.02	0.24	0.46	0.36
		1	0.85	0.01	0.06	0	4.29	0.3	0.53	0.34	0.85	0.01	0.06	0	4.34	0.28	0.52	0.32	**0.86**	**0.01**	**0.06**	**0**	**4.33**	**0.29**	**0.55**	**0.43**
		2	0.8	0.07	0.07	0.01	5.45	0.79	0.75	0.66	0.82	0.04	0.07	0.01	5.32	0.55	0.57	0.55	0.84	0.03	0.07	0.01	5.16	0.47	0.65	0.49
		3	0.31	0.54	0.13	0.04	9.12	2.85	3.41	3.33	0.41	0.39	0.13	0.03	8.89	2.42	2.81	2.09	0.57	0.23	0.11	0.02	7.92	1.64	1.99	1.38
		4	0.39	0.56	0.13	0.04	9.42	2.79	3.51	2.92	0.56	0.28	0.12	0.02	8.59	1.78	2.58	2.15	0.58	0.54	0.12	0.03	8.29	2.4	2.82	2.34
	2	0	0.82	0.01	0.06	0	4.34	0.25	0.6	0.39	0.83	0.01	0.06	0	4.32	0.29	0.61	0.4	0.83	0.01	0.06	0	4.3	0.25	0.66	0.43
		1	0.84	0.01	0.06	0	4.43	0.35	0.64	0.48	0.84	0.01	0.06	0	4.47	0.34	0.66	0.46	**0.85**	**0.01**	**0.06**	**0**	**4.42**	**0.35**	**0.66**	**0.46**
		2	0.8	0.04	0.08	0.01	5.48	0.65	0.95	0.85	0.82	0.04	0.08	0.01	5.33	0.73	0.65	0.52	0.84	0.02	0.07	0.01	5.11	0.47	0.73	0.45
		3	0.4	0.36	0.12	0.03	8.7	2.08	2.96	2.08	0.52	0.3	0.12	0.03	8.16	1.75	2.4	2.24	0.53	0.31	0.11	0.03	8.03	2.17	2.29	2.37
		4	0.44	0.36	0.13	0.03	9.55	2.35	3.36	2.34	0.63	0.21	0.12	0.03	8.31	1.82	2.56	1.59	0.69	0.19	0.11	0.03	8.03	1.85	2.33	2.21
	3	0	0.81	0.01	0.06	0.01	4.38	0.37	0.57	0.43	0.81	0.01	0.06	0	4.39	0.33	0.55	0.39	0.82	0.01	0.06	0	4.35	0.32	0.56	0.4
		1	0.83	0.02	0.06	0	4.48	0.23	0.52	0.38	0.84	0.01	0.06	0.01	4.49	0.29	0.52	0.41	**0.84**	**0.01**	**0.06**	**0**	**4.5**	**0.29**	**0.5**	**0.44**
		2	0.76	0.08	0.08	0.01	5.77	0.77	0.72	0.54	0.77	0.11	0.08	0.02	5.77	1.14	0.73	0.49	0.81	0.03	0.08	0.01	5.51	0.5	0.85	0.62
		3	0.34	0.31	0.14	0.03	9.72	2.31	3.29	2.64	0.47	0.28	0.12	0.03	8.7	1.88	2.54	2.25	0.56	0.24	0.12	0.02	8.2	1.66	2.19	1.77
		4	0.15	0.41	0.14	0.03	25.4	5.38	3.4	2.47	0.27	0.44	0.13	0.02	24.1	4.99	3.36	2.42	0.38	0.35	0.12	0.02	23.2	4.71	2.94	2.23
	4	0	0.8	0.01	0.07	0	4.63	0.29	0.76	0.52	0.8	0.01	0.07	0	4.58	0.28	0.74	0.54	**0.81**	**0.01**	**0.07**	**0**	**4.54**	**0.21**	**0.66**	**0.54**
		1	0.81	0.02	0.07	0	4.88	0.31	0.67	0.52	0.82	0.01	0.07	0	4.81	0.33	0.66	0.57	0.83	0.01	0.07	0.01	4.8	0.34	0.74	0.6
		2	0.77	0.05	0.08	0.01	5.92	0.66	1	0.76	0.78	0.04	0.08	0.01	5.83	0.54	0.81	0.65	0.8	0.05	0.08	0.01	5.8	0.67	0.87	0.72
		3	0.32	0.42	0.13	0.04	9.5	2.42	3.16	2.4	0.34	0.31	0.14	0.03	9.56	2.09	3	2.39	0.52	0.25	0.12	0.03	8.51	1.8	2.38	2.13
		4	0.42	0.5	0.13	0.04	9.36	2.54	2.5	2.34	0.57	0.3	0.12	0.03	8.73	2	2.4	2.27	0.71	0.2	0.11	0.02	7.94	1.7	1.94	1.58
	5	0	0.77	0.01	0.07	0	4.97	0.32	0.82	0.44	0.78	0.01	0.07	0.01	4.87	0.34	0.8	0.46	0.79	0.01	0.07	0.01	4.87	0.34	0.75	0.46
		1	0.8	0.01	0.07	0	4.96	0.35	0.65	0.52	0.81	0.02	0.07	0.01	4.9	0.35	0.65	0.55	**0.82**	**0.01**	**0.07**	**0**	**4.87**	**0.33**	**0.71**	**0.51**
		2	0.7	0.1	0.09	0.01	6.57	1.07	1.12	0.83	0.74	0.09	0.09	0.01	6.3	1.01	0.99	0.56	0.78	0.05	0.09	0.01	6.03	0.73	1.05	0.84
		3	0.3	0.52	0.13	0.04	9.49	2.86	3.29	2.95	0.41	0.34	0.12	0.03	8.82	2.04	3.06	2.36	0.51	0.28	0.12	0.03	8.5	1.85	2.4	1.88
		4	0.35	0.56	0.14	0.04	9.85	2.85	3.38	2.95	0.57	0.23	0.12	0.03	8.85	1.8	2.79	1.85	0.68	0.24	0.11	0.02	8.06	1.79	2.3	2.03

Para la ciudad de Quito, tenemos la tabla de resultados donde podemos observar que el número de retrasos marca una diferencia con respecto a la bondad de ajuste y las métricas de validación. Sin embargo, el número de neuronas no afecta mucho ya que no hay mucha diferencia entre los valores presentados. Los valores más destacados para análisis se pueden observar que la bondad de ajuste es más alta cuando el pronóstico es más cercano caso contrario cuando el día a pronosticar se aleja tanto la bondad de ajuste disminuye y el error tiende a aumentar. Por ejemplo para el día 1 tiene un R^2=0,86, MAPE=0,06, RMSE=4,33 y PBIAS=0,55 mostrando los mejores resultados y para el día 5 podemos ver que tiene un R^2=0,82, MAPE=0,07, RMSE=4,87 y PBIAS=0,71mostrando los resultados más bajos en comparación con los demás días.

6.4.1.2. Predicción del NO_2 mediante Redes Neuronales Recurrentes.

Las siguientes figuras muestran el progreso del entrenamiento de la red en función de las épocas, para la predicción de NO_2 para 1 día hacia adelante. La arquitectura de la red utilizada fue: 1 capa oculta, 64 neuronas, y un número de retrasos de 2.

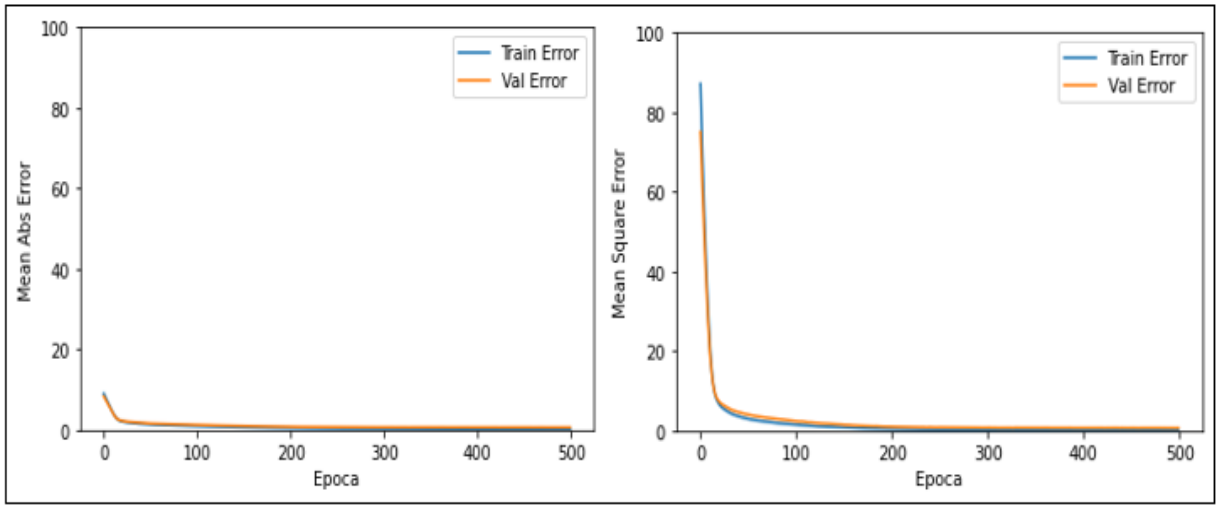

Gráfico 42. Entrenamiento de la red neuronal en función de las épocas para el NO_2.

Para el caso del NO_2 podemos apreciar que conforme avanza el entrenamiento, el error en el conjunto de entrenamiento sigue decreciendo. Además, hay que tener en cuenta que no existe mucha diferencia entre el error en el conjunto de entrenamiento y el error en el conjunto de validación son muy similares. Para este caso también se aplicó la función *EarlyStopping* para saber cuándo se detiene el entrenamiento. Las siguientes figuras muestran el avance y finalización del entrenamiento.

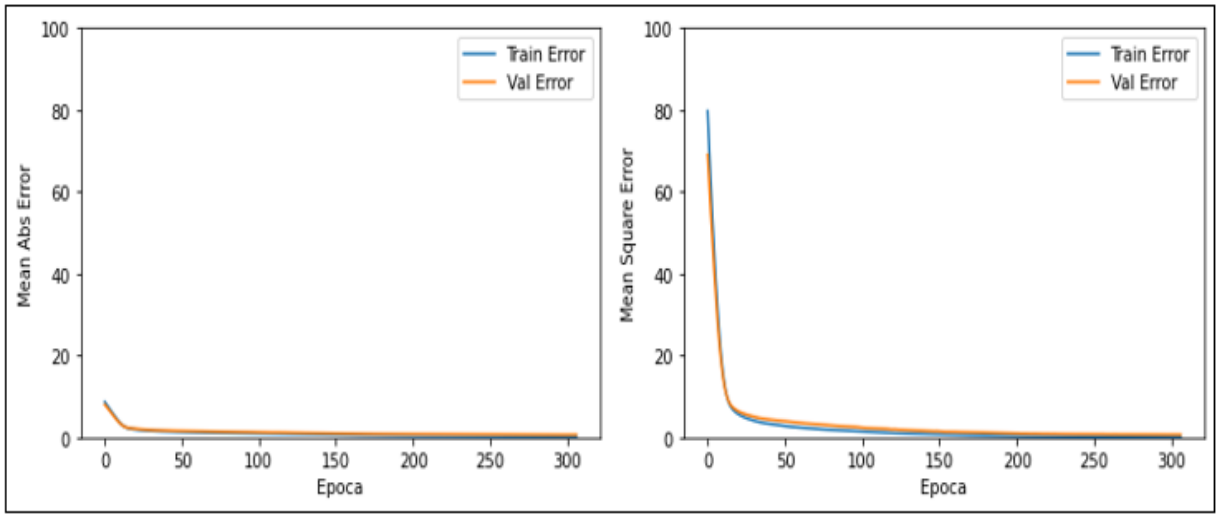

Gráfico 43. Entrenamiento de la red neuronal aplicando detección temprana del error de validación del NO_2.

Estas figuras muestran que el entrenamiento se detuvo en la época 300 y que en este punto se obtuvo la red con el entrenamiento óptimo.

El siguiente gráfico muestra la relación entre los valores de NO_2 verdaderos y los valores predichos por la red entrenada para el conjunto de prueba.

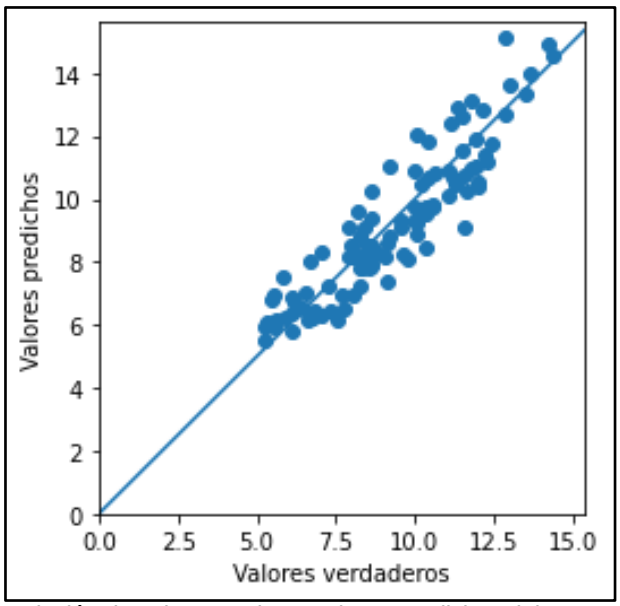

Gráfico 44. Gráfico de relación de valores reales y valores predichos del NO_2 para Redes Neuronales Recurrentes.

En un ajuste perfecto, los puntos caerían sobre la línea azul. Analizamos los residuos entre el valor verdadero y el valor predicho por la red.

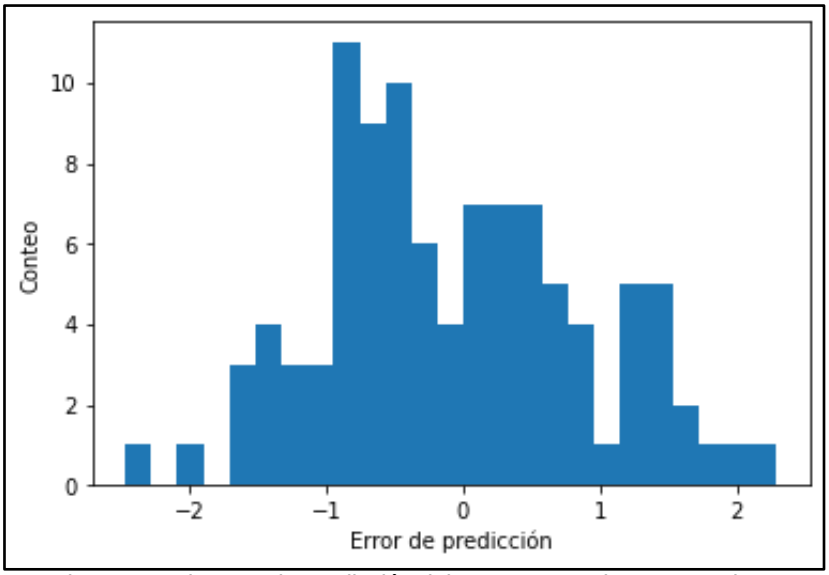

Gráfico 45. Histograma de error de predicción del NO_2 para Redes Neuronales Recurrentes.

Los valores de los errores tienden a seguir una distribución normal y el valor central tiende a 0. Este es un indicador de que el modelo está capturando algún patrón (si lo hay) entre las variables independiente y la dependiente.

De la misma manera, para este contaminante, el proceso anterior se repitió 30 veces con una inicialización diferente de los pesos de la red y con una partición diferente del conjunto inicial para el entrenamiento y la prueba. Las siguientes tablas muestran los resultados del promedio y desviación estándar de las métricas de evaluación de bondad de ajuste y de error.

Tabla 14. Resultados de análisis de Redes Neuronales Recurrentes para NO_2 de Cuenca.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.79	0.01	0.09	0.01	0.54	0.04	0.9	0.58	0.8	0.01	0.09	0.01	0.54	0.04	1.03	0.72	**0.81**	**0.01**	**0.09**	**0.01**	**0.54**	**0.04**	**0.98**	**0.74**
		1	0.79	0.03	0.1	0.01	0.61	0.05	1.27	0.94	0.8	0.03	0.1	0.01	0.61	0.05	1.13	0.88	0.83	0.02	0.1	0.01	0.6	0.04	1.05	0.8
		2	0.76	0.07	0.12	0.01	0.71	0.06	1.34	0.73	0.77	0.07	0.12	0.01	0.71	0.07	1	0.67	0.81	0.06	0.12	0.01	0.7	0.06	1.17	0.92
		3	0.75	0.07	0.14	0.01	0.8	0.07	1.02	2.14	0.78	0.07	0.14	0.01	0.81	0.07	1.05	2.19	0.8	0.07	0.13	0.01	0.79	0.06	0.65	2.01
		4	0.69	0.1	0.16	0.02	0.93	0.1	2.14	1.41	0.74	0.07	0.15	0.01	0.9	0.07	1.66	1.47	0.8	0.04	0.15	0.01	0.87	0.07	1.62	1.27
	2	0	0.76	0.02	0.1	0.01	0.59	0.04	1.28	0.76	0.77	0.02	0.1	0.01	0.58	0.04	1.19	0.84	0.78	0.01	0.1	0.01	0.58	0.04	1.16	0.81
		1	0.76	0.03	0.11	0.01	0.64	0.06	1.16	0.84	0.77	0.06	0.11	0.01	0.65	0.07	0.98	0.83	**0.82**	**0.03**	**0.11**	**0.01**	**0.63**	**0.04**	**0.97**	**0.93**
		2	0.75	0.06	0.12	0.01	0.72	0.06	1.22	0.7	0.78	0.05	0.12	0.01	0.71	0.05	1.04	0.71	0.79	0.06	0.12	0.01	0.73	0.06	1.02	0.77
		3	0.73	0.09	0.14	0.02	0.84	0.1	1.66	1.42	0.75	0.07	0.14	0.01	0.82	0.07	1.54	1.11	0.77	0.06	0.14	0.01	0.83	0.07	1.44	1.02
		4	0.69	0.1	0.16	0.02	0.92	0.08	2.53	1.81	0.77	0.06	0.15	0.01	0.9	0.07	1.77	1.22	0.8	0.06	0.15	0.01	0.89	0.08	1.62	1.57
	3	0	0.74	0.02	0.1	0.01	0.61	0.05	0.9	0.61	0.75	0.02	0.1	0.01	0.61	0.05	0.84	0.53	**0.77**	**0.02**	**0.1**	**0.01**	**0.59**	**0.04**	**0.81**	**0.65**
		1	0.74	0.02	0.11	0.01	0.67	0.05	1.12	0.96	0.74	0.06	0.12	0.01	0.69	0.06	1.26	1	0.78	0.04	0.11	0.01	0.67	0.06	1.3	0.93
		2	0.72	0.07	0.13	0.01	0.77	0.07	1.51	1.16	0.74	0.05	0.13	0.01	0.79	0.06	1.3	1.12	0.77	0.07	0.13	0.01	0.78	0.07	1.29	1.29
		3	0.72	0.07	0.15	0.01	0.88	0.08	1.78	1.31	0.74	0.07	0.15	0.02	0.86	0.08	1.65	1.08	0.78	0.07	0.14	0.01	0.85	0.07	1.75	1.14
		4	0.72	0.07	0.16	0.01	0.95	0.08	2.45	2.26	0.74	0.07	0.16	0.01	0.93	0.08	1.96	1.38	0.8	0.03	0.15	0.01	0.89	0.07	2.1	1.51
	4	0	0.71	0.02	0.11	0.01	0.65	0.04	1.04	0.67	0.73	0.02	0.11	0.01	0.64	0.05	0.99	0.76	0.74	0.02	0.11	0.01	0.64	0.04	0.99	0.8
		1	0.72	0.04	0.12	0.01	0.7	0.06	1.88	1.24	0.73	0.07	0.12	0.02	0.71	0.08	1.76	1.09	**0.77**	**0.04**	**0.12**	**0.01**	**0.69**	**0.05**	**1.47**	**1.38**
		2	0.71	0.08	0.13	0.01	0.79	0.07	1.54	1.45	0.72	0.1	0.13	0.01	0.78	0.08	1.63	1.4	0.78	0.06	0.13	0.01	0.78	0.05	1.48	1.13
		3	0.67	0.1	0.15	0.02	0.88	0.09	1.83	1.19	0.75	0.05	0.14	0.01	0.83	0.06	1.95	1.23	0.78	0.06	0.14	0.02	0.85	0.08	1.44	1.14
		4	0.67	0.09	0.17	0.02	0.97	0.1	1.98	1.34	0.74	0.08	0.16	0.02	0.94	0.1	1.78	1.35	0.74	0.08	0.16	0.01	0.94	0.07	1.8	1.56
	5	0	0.7	0.02	0.11	0.01	0.65	0.04	0.65	0.04	0.72	0.02	0.11	0.01	0.65	0.04	0.95	0.92	0.73	0.02	0.11	0.01	0.64	0.03	0.92	0.73
		1	0.69	0.05	0.12	0.01	0.71	0.05	1.5	1.13	0.73	0.09	0.12	0.01	0.69	0.08	1.5	1.02	**0.77**	**0.04**	**0.11**	**0.01**	**0.67**	**0.05**	**1.5**	**1.15**
		2	0.68	0.1	0.13	0.01	0.78	0.08	1.84	1.16	0.73	0.07	0.13	0.01	0.79	0.07	1.63	1.27	0.77	0.07	0.13	0.01	0.76	0.07	1.57	1.33
		3	0.7	0.1	0.15	0.01	0.88	0.07	2.01	1.4	0.75	0.07	0.14	0.01	0.85	0.06	1.66	1.4	0.78	0.08	0.14	0.01	0.84	0.06	1.65	1.05
		4	0.7	0.1	0.16	0.01	0.93	0.09	2.04	1.63	0.76	0.05	0.15	0.01	0.91	0.06	2.21	1.72	0.78	0.06	0.15	0.01	0.89	0.08	2.02	1.65

Podemos observar en la tabla que el modelo nos da mejores resultados cuando aumenta el número de neuronas, esto da significado puesto que para todos los días predichos los valores más certeros tomando en cuenta las métricas de validación se dan para las sesenta y cuatro neuronas. Además, cuando consideramos más variables “ficticias” es decir las variables creadas con retrasos, el rendimiento general no mejora significativamente. Finalmente se puede observar que mientras se trata de predecir un día más alejado, la bondad de ajuste de los modelos tiende a descender. Los mejores resultados en cuanto a un balance de las métricas de bondad de ajuste y de error se presentan en resaltadas en negrita. Para la ciudad de Cuenca los mejores valores pronóstico en 1 día fueron de R^2=0,81, MAPE=0,09, RMSE=0,54, PBIAS=0,98. Para el pronóstico en dos días los valores

fueron de R^2=0,82, MAPE=0,11, RMSE=0,63, PBIAS=0,97. Para el pronóstico en tres días los valores fueron de R^2=0,77, MAPE=0,1, RMSE=0,59, PBIAS=0,81. Para el pronóstico en cuatro días los valores fueron de R^2=0,77, MAPE=0,12, RMSE=0,69, PBIAS=1,47. Para el pronóstico en cinco días los valores fueron de R^2=0,77, MAPE=0,11, RMSE=0,67, PBIAS=1,5.

Tabla 15. Resultados de análisis de Redes Neuronales Recurrentes para NO_2 de Guayaquil.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.79	0.01	0.07	0.01	0.76	0.05	0.76	0.54	0.8	0.01	0.07	0.01	0.76	0.06	0.71	0.43	0.8	0.01	0.07	0.01	0.76	0.05	0.71	0.46
		1	0.81	0.03	0.07	0.01	0.79	0.08	0.98	0.72	0.83	0.02	0.07	0.01	0.79	0.05	0.9	0.78	**0.85**	**0.02**	**0.07**	**0.01**	**0.79**	**0.08**	**0.81**	**0.79**
		2	0.78	0.13	0.09	0.02	0.98	0.18	1.25	0.88	0.83	0.03	0.09	0.01	0.96	0.08	1.06	0.82	0.86	0.02	0.09	0.01	0.92	0.07	1.05	0.86
		3	0.77	0.09	0.1	0.01	1.08	0.14	1.39	0.88	0.82	0.05	0.1	0.01	1.07	0.09	1.47	1	0.81	0.15	0.1	0.01	1.11	0.16	1.18	1
		4	0.73	0.15	0.12	0.02	1.31	0.24	2.15	1.73	0.8	0.05	0.12	0.01	1.26	0.15	1.72	1.52	0.81	0.04	0.12	0.01	1.27	0.12	1.41	1.29
	2	0	0.74	0.02	0.08	0.01	0.85	0.07	0.95	0.63	0.76	0.02	0.08	0.01	0.84	0.07	0.84	0.6	0.76	0.02	0.08	0.01	0.84	0.07	0.87	0.52
		1	0.78	0.02	0.08	0.01	0.86	0.07	0.87	0.65	0.8	0.02	0.08	0.01	0.86	0.07	1.07	0.72	**0.83**	**0.01**	**0.08**	**0.01**	**0.86**	**0.08**	**0.95**	**0.66**
		2	0.76	0.06	0.09	0.01	0.99	0.11	1.05	0.79	0.8	0.03	0.09	0.01	1	0.11	1.29	1.14	0.83	0.03	0.09	0.01	0.97	0.09	0.96	0.87
		3	0.77	0.07	0.1	0.01	1.11	0.12	1.43	1.12	0.76	0.12	0.11	0.02	1.18	0.19	1.59	1.05	0.82	0.06	0.11	0.01	1.16	0.12	1.47	1.04
		4	0.72	0.14	0.12	0.02	1.32	0.2	1.95	1.56	0.76	0.07	0.12	0.01	1.33	0.16	1.55	1	0.79	0.05	0.12	0.01	1.32	0.14	1.61	1.09
	3	0	0.72	0.02	0.08	0	0.89	0.05	0.85	0.63	0.74	0.02	0.08	0.01	0.89	0.06	0.96	0.56	0.75	0.02	0.08	0.01	0.88	0.06	0.94	0.61
		1	0.74	0.02	0.09	0.01	0.96	0.06	1.13	0.84	0.77	0.03	0.09	0.01	0.95	0.06	1.11	0.76	**0.79**	**0.03**	**0.09**	**0.01**	**0.94**	**0.05**	**1.09**	**0.79**
		2	0.68	0.15	0.11	0.02	1.14	0.18	1.52	1.33	0.73	0.09	0.1	0.01	1.12	0.13	1.32	1.09	0.77	0.11	0.1	0.01	1.12	0.14	1.23	1.13
		3	0.72	0.11	0.11	0.02	1.23	0.18	1.18	0.89	0.78	0.07	0.11	0.01	1.18	0.13	1.42	1	0.81	0.05	0.11	0.01	1.18	0.11	1.44	0.96
		4	0.71	0.09	0.13	0.02	1.43	0.16	2.43	1.39	0.76	0.08	0.13	0.01	1.38	0.16	2.14	1.35	0.79	0.07	0.13	0.01	1.35	0.14	1.89	1.37
	4	0	0.69	0.02	0.09	0.01	0.93	0.06	0.85	0.57	0.71	0.01	0.09	0.01	0.93	0.06	0.81	0.54	0.72	0.01	0.09	0.01	0.92	0.06	0.78	0.56
		1	0.7	0.03	0.09	0.01	0.98	0.06	1	0.7	0.74	0.02	0.09	0.01	0.98	0.07	0.89	0.78	**0.77**	**0.03**	**0.09**	**0.01**	**0.96**	**0.06**	**0.84**	**0.62**
		2	0.71	0.05	0.1	0.01	1.12	0.12	0.91	0.6	0.75	0.06	0.1	0.01	1.1	0.13	1.3	0.87	0.78	0.05	0.1	0.01	1.1	0.11	1.06	0.81
		3	0.61	0.21	0.13	0.02	1.35	0.23	2.2	1.61	0.75	0.07	0.12	0.01	1.24	0.11	1.47	1.06	0.79	0.06	0.12	0.01	1.25	0.13	1.5	1.25
		4	0.61	0.3	0.14	0.03	1.49	0.25	2.17	1.92	0.71	0.19	0.13	0.02	1.4	0.18	2.08	1.32	0.77	0.09	0.13	0.01	1.39	0.13	1.9	1.23
	5	0	0.66	0.02	0.09	0.01	0.96	0.05	0.77	0.68	0.68	0.02	0.09	0.01	0.95	0.05	0.82	0.59	0.69	0.02	0.09	0.01	0.94	0.05	0.7	0.77
		1	0.68	0.03	0.1	0.01	1	0.09	1.1	0.86	0.71	0.03	0.09	0.01	1	0.08	1.16	0.85	**0.75**	**0.03**	**0.09**	**0.01**	**0.98**	**0.07**	**1.14**	**0.83**
		2	0.69	0.09	0.11	0.01	1.15	0.12	1.33	0.94	0.72	0.05	0.11	0.01	1.15	0.09	1.4	0.95	0.77	0.07	0.11	0.01	1.13	0.1	1.14	0.86
		3	0.68	0.11	0.12	0.01	1.26	0.13	1.59	1.3	0.71	0.09	0.12	0.01	1.25	0.11	1.56	1.23	0.78	0.05	0.12	0.01	1.24	0.12	1.38	1.14
		4	0.66	0.14	0.13	0.02	1.42	0.15	2.36	1.77	0.76	0.07	0.13	0.01	1.35	0.13	1.75	1.38	0.77	0.07	0.13	0.01	1.36	0.15	1.65	1.36

De igual manera, para la ciudad de Guayaquil se nota que un número de neuronas alto representa mejores resultados tomando en cuenta de igual manera las métricas de validación y la bondad de ajuste. Además, cuando consideramos las variables “ficticias” es decir las variables creadas con retrasos, podemos observar que para todos los días pronosticados el número de retrasos adecuado es el de 2. Los mejores valores pronóstico de 1 día fueron de R^2=0,85, MAPE=0,07, RMSE=0,79, PBIAS=0,81. Los mejores valores pronosticados en 2 días fueron de R^2=0,83, MAPE=0,08, RMSE=0,86, PBIAS=0,95. Los mejores valores pronosticados en 3 días fueron de R^2=0,79, MAPE=0,09, RMSE=0,94, PBIAS=1,09. Los mejores valores pronosticados en 4 días fueron de R^2=0,77, MAPE=0,09, RMSE=0,96, PBIAS=0,84. Los mejores valores pronosticados en 5 días fueron de R^2=0,75, MAPE=0,09, RMSE=0,98, PBIAS=1,14. De Igual forma el pronóstico para días más alejados representa valores inadecuados que los primeros días. Los mejores resultados en cuanto a un balance de las métricas de bondad de ajuste y de error se presentan en resaltadas en negrita.

Tabla 16. Resultados de análisis de Redes Neuronales Recurrentes para NO_2 de Quito.

			16 Neuronas								32 Neuronas								64 Neuronas							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.78	0.01	0.07	0.01	0.82	0.04	0.74	0.64	0.79	0.01	0.07	0.01	0.82	0.04	0.76	0.6	**0.8**	**0.01**	**0.07**	**0.01**	**0.81**	**0.04**	**0.72**	**0.56**
		1	0.78	0.02	0.08	0.01	0.91	0.06	1.03	0.67	0.79	0.03	0.08	0.01	0.9	0.07	0.83	0.69	0.82	0.02	0.08	0.01	0.89	0.06	1.07	0.7
		2	0.81	0.03	0.1	0.01	1.04	0.07	1.13	0.84	0.83	0.03	0.1	0.01	1.04	0.06	0.93	0.68	0.86	0.03	0.1	0.01	1.06	0.07	1.01	0.9
		3	0.79	0.07	0.11	0.01	1.21	0.12	1.71	1.24	0.81	0.03	0.11	0.01	1.22	0.07	1.42	0.84	0.84	0.03	0.11	0.01	1.21	0.09	1.27	0.85
		4	0.71	0.17	0.13	0.02	1.44	0.17	2.07	1.45	0.77	0.09	0.14	0.01	1.46	0.13	1.88	1.55	0.79	0.07	0.13	0.01	1.43	0.11	1.94	1.43
	2	0	0.71	0.01	0.09	0.01	0.96	0.07	0.95	0.67	0.72	0.02	0.09	0.01	0.96	0.06	1.04	0.61	0.73	0.02	0.09	0.01	0.95	0.07	1.11	0.68
		1	0.7	0.03	0.09	0.01	1.03	0.07	0.93	0.81	0.72	0.02	0.09	0.01	1.03	0.07	0.98	0.77	**0.74**	**0.03**	**0.09**	**0.01**	**1.02**	**0.07**	**0.92**	**0.8**
		2	0.75	0.03	0.11	0.01	1.16	0.1	1.32	1.16	0.78	0.03	0.11	0.01	1.15	0.09	1.21	1.07	0.8	0.04	0.11	0.01	1.18	0.09	1.16	1.1
		3	0.73	0.05	0.12	0.01	1.34	0.14	1.7	1.38	0.76	0.07	0.12	0.01	1.33	0.13	1.49	1.16	0.78	0.05	0.12	0.01	1.34	0.11	1.45	1.44
		4	0.63	0.19	0.15	0.02	1.6	0.2	1.92	1.51	0.69	0.16	0.14	0.02	1.55	0.18	1.81	1.79	0.75	0.08	0.14	0.01	1.49	0.14	2	1.47
	3	0	0.66	0.02	0.1	0.01	1.04	0.05	1.07	0.77	0.68	0.02	0.1	0.01	1.05	0.07	1.05	0.84	0.69	0.02	0.1	0.01	1.03	0.06	1.04	0.92
		1	0.65	0.04	0.1	0.01	1.12	0.06	1.04	0.97	0.68	0.04	0.1	0.01	1.1	0.07	1.12	0.88	0.7	0.04	0.1	0.01	1.09	0.06	1.11	0.8
		2	0.71	0.04	0.11	0.01	1.2	0.1	1.19	0.8	0.75	0.04	0.11	0.01	1.18	0.09	1.1	0.68	**0.77**	**0.04**	**0.11**	**0.01**	**1.16**	**0.09**	**1.06**	**0.78**
		3	0.71	0.05	0.12	0.01	1.35	0.14	1.9	1.47	0.74	0.05	0.12	0.01	1.35	0.11	1.64	1.21	0.76	0.06	0.13	0.01	1.36	0.12	1.67	1.47
		4	0.63	0.12	0.15	0.02	1.6	0.2	2.42	1.93	0.68	0.13	0.14	0.01	1.57	0.12	2.36	1.46	0.73	0.09	0.14	0.01	1.54	0.11	2.35	1.45
	4	0	0.65	0.02	0.1	0.01	1.03	0.06	1.11	0.63	0.67	0.02	0.1	0.01	1.02	0.06	1.12	0.76	0.68	0.01	0.1	0.01	1.01	0.05	1.05	0.77
		1	0.66	0.03	0.1	0.01	1.11	0.09	1.24	0.87	0.68	0.03	0.1	0.01	1.1	0.06	1.21	0.84	0.7	0.03	0.1	0.01	1.1	0.06	1.25	0.89
		2	0.71	0.03	0.11	0.01	1.19	0.08	1.35	1	0.75	0.03	0.11	0.01	1.17	0.08	1.39	0.82	**0.78**	**0.03**	**0.11**	**0.01**	**1.17**	**0.07**	**1.21**	**0.87**
		3	0.71	0.06	0.13	0.01	1.36	0.11	1.75	1.25	0.77	0.04	0.12	0.01	1.34	0.12	1.58	1.21	0.77	0.11	0.13	0.01	1.37	0.13	1.66	1.48
		4	0.6	0.21	0.14	0.02	1.54	0.19	1.95	1.76	0.68	0.15	0.14	0.02	1.56	0.16	1.61	1.46	0.7	0.18	0.14	0.01	1.49	0.15	1.46	1.23
	5	0	0.66	0.02	0.09	0.01	1.01	0.07	0.82	0.7	0.67	0.02	0.09	0.01	1.02	0.07	0.74	0.64	0.68	0.01	0.09	0.01	1.01	0.06	0.8	0.7
		1	0.65	0.05	0.1	0.01	1.11	0.07	1.03	0.79	0.67	0.07	0.1	0.01	1.12	0.07	1.07	0.79	0.69	0.05	0.1	0.01	1.1	0.08	1.02	0.8
		2	0.73	0.03	0.11	0.01	1.18	0.08	0.95	0.86	0.76	0.03	0.11	0.01	1.18	0.09	1.16	0.91	**0.78**	**0.03**	**0.11**	**0.01**	**1.18**	**0.08**	**1.22**	**0.88**
		3	0.71	0.16	0.13	0.01	1.37	0.15	1.54	1.03	0.76	0.11	0.13	0.01	1.38	0.14	1.59	1.26	0.8	0.04	0.13	0.01	1.34	0.1	1.43	0.92
		4	0.67	0.13	0.14	0.02	1.53	0.19	1.86	1.39	0.7	0.17	0.14	0.02	1.51	0.18	2.02	1.26	0.74	0.17	0.14	0.02	1.5	0.16	1.84	1.29

La ciudad de Quito sigue el mismo patrón que se presentó en las dos otras ciudades. Siendo que un número de neuronas alto representa mejores resultados. En este caso cuando consideramos las variables “ficticias” es decir las variables creadas con retazos, podemos observar que para todos los días pronosticados el número de retrasos adecuado va desde 3 hasta 1, siendo los peores los valores con retrasos de 4 y 5. Los mejores valores pronóstico de 1 día fueron de R^2=0,8, MAPE=0,07, RMSE=0,81, PBIAS=0,72. Los mejores valores pronosticados en 2 días fueron de R^2=0,74, MAPE=0,09, RMSE=1,02, PBIAS=0,92. Los mejores valores pronosticados en 3 días fueron de R^2=0,77, MAPE=0,11, RMSE=1,16, PBIAS=1,06. Los mejores valores pronosticados en 4 días fueron de R^2=0,78, MAPE=0,11, RMSE=1,17, PBIAS=1,21. Los mejores valores pronosticados en 5 días fueron de R^2=0,78, MAPE=0,11, RMSE=1,18, PBIAS=1,22. De Igual forma el pronóstico para días más alejados representa valores menos adecuados y los mejores resultados en cuanto a un balance de las métricas de bondad de ajuste y de error se presentan en resaltadas en negrita.

6.5. Random Forest

Para este algoritmo no fue necesario normalizar las variables independientes. Se utilizó la librería sklearn con los hiperparámetros definidos por defecto. Todo fue realizado en el entorno Google Colab que usa el lenguaje Python. Cabe destacar que para este método se utilizó dos diseños experimentales donde el primero se tomó como característica principal la profundidad máxima el segundo tomó en cuenta el número mínimo de muestras.

Para las 3 ciudades se entrenaron varias configuraciones de bosques aleatorios siguiendo el siguiente diseño experimental.

Tabla 17. Configuraciones de Bosques Aleatorios.

Número de árboles	*100*
Función de medida de calidad	Error cuadrado
Número de repeticiones de cada configuración	30
Profundidad máxima	5, 15, 25
Mínimo número de muestras en los nodos hoja	5, 15, 25
Número de retrasos	0, 1, 2, 3, 4
Días hacia adelante para la predicción	1, 2, 3, 4, 5

Para el caso de las concentraciones de gases de SO_2 se puede apreciar en las siguientes figuras la tendencia de los resultados de los valores reales y los valores predichos, donde se observa que los valores se posicionan más cerca de la línea de tendencia lo que indica un mejor ajuste del modelo.

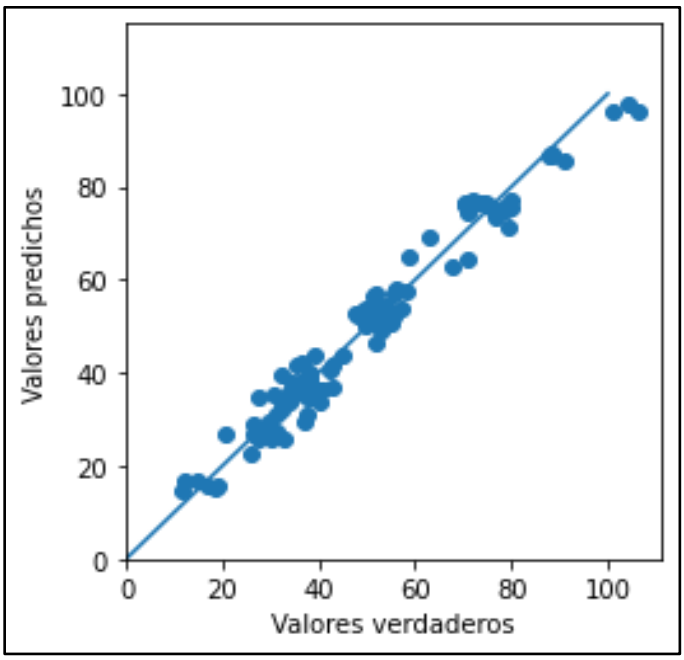

Gráfico 46. Relación de valores reales y valores predichos del SO_2 para Random Forest.

Con respecto a los errores mostrados en el siguiente histograma se puede observar que muestra una distribución normal donde su punto central tiende a 0 lo que indica que existe algún patrón entre las variables estudiadas.

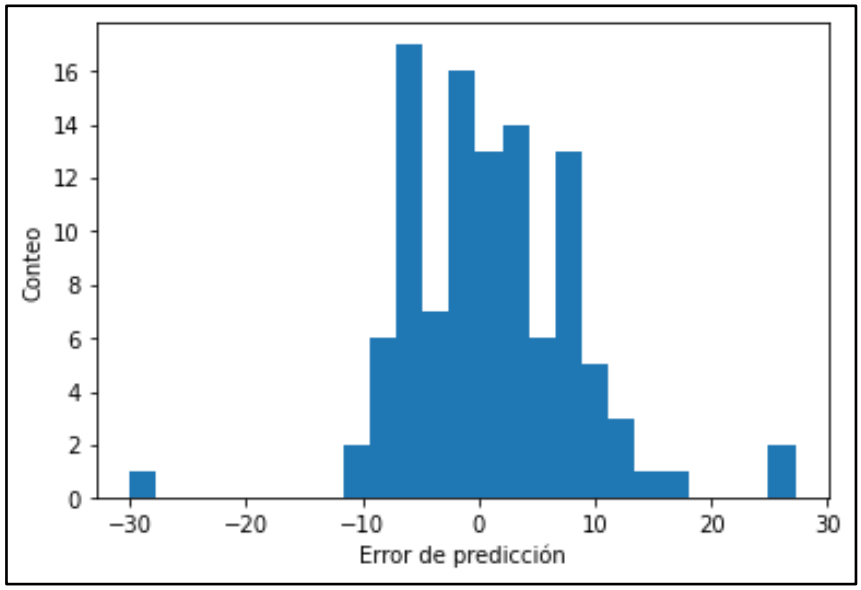

Gráfico 47. Histograma de error de predicción del SO_2 para Random Forest.

Cuando se refiere al NO_2, se puede observar en la tabla la relación de los resultados de los valores reales con los valores predichos donde muestra cercanía de muchos de los

datos hacia la línea de tendencia. sin embargo también existen valores que se encuentran más alejados de dicha línea lo que muestra que no existe mucha relación con todos los datos.

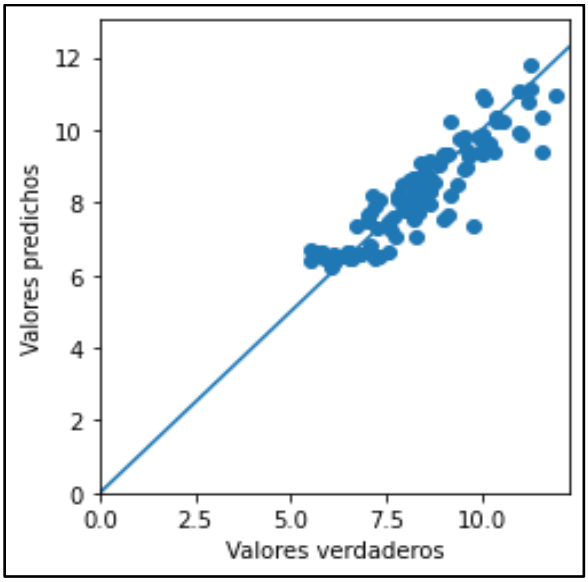

Gráfico 48. Gráfico de relación de valores reales y valores predichos del NO_2 para Random Forest.

Los valores de los errores de predicción tienden a seguir una distribución normal. En este caso podemos observar que el valor central tiende a 0 variando entre 1 y -1, esto quiere decir que el modelo está capturando patrones para la predicción.

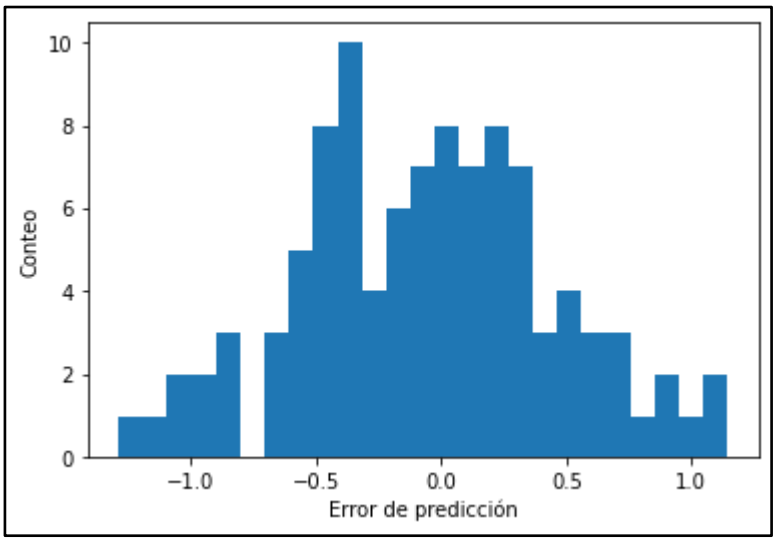

Gráfico 49. Histograma de error de predicción del NO_2 para Random Forest.

6.5.1.1. Predicción del SO_2 mediante Random Forest considerando la Profundidad máxima de cada árbol de decisión.

Aquí se tomó como característica principal la profundidad máxima de cada árbol de decisión donde tenemos los siguientes resultados.

Tabla 18. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para SO_2 de Cuenca.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.92	0	0.06	0	3.44	0.22	0.53	0.4	0.98	0	0.06	0.01	3.51	0.23	0.55	0.45	0.98	0	0.06	0.01	3.51	0.22	0.55	0.45
		1	0.94	0	0.06	0	3.4	0.2	0.64	0.38	0.98	0	0.06	0	3.42	0.2	0.62	0.38	0.98	0	0.06	0	3.41	0.2	0.62	0.38
		2	0.95	0	0.06	0	3.28	0.2	0.47	0.49	**0.98**	**0**	**0.06**	**0**	**3.3**	**0.21**	**0.43**	**0.47**	0.98	0	0.06	0	3.3	0.21	0.44	0.47
		3	0.95	0	0.06	0	3.2	0.2	0.71	0.45	0.98	0	0.06	0.01	3.22	0.21	0.72	0.43	0.98	0	0.06	0.01	3.22	0.21	0.72	0.43
		4	0.95	0	0.06	0	3.26	0.15	0.51	0.41	0.98	0	0.06	0	3.27	0.17	0.51	0.39	0.98	0	0.06	0	3.27	0.17	0.49	0.39
	2	0	0.91	0	0.07	0.01	3.6	0.24	0.58	0.4	0.98	0	0.07	0.01	3.63	0.25	0.55	0.4	0.98	0	0.07	0.01	3.63	0.25	0.55	0.4
		1	0.93	0	0.07	0.01	3.54	0.26	0.73	0.65	0.98	0	0.07	0.01	3.52	0.26	0.69	0.67	0.98	0	0.07	0.01	3.52	0.26	0.69	0.66
		2	0.94	0	0.07	0	3.54	0.17	0.69	0.62	0.98	0	0.07	0	3.53	0.16	0.68	0.63	0.98	0	0.07	0	3.53	0.16	0.67	0.64
		3	0.94	0	0.07	0.01	3.45	0.22	0.7	0.53	**0.98**	**0**	**0.07**	**0.01**	**3.43**	**0.23**	**0.72**	**0.51**	0.98	0	0.07	0.01	3.43	0.23	0.73	0.52
		4	0.94	0	0.07	0.01	3.54	0.23	0.81	0.48	0.98	0	0.07	0.01	3.54	0.23	0.83	0.49	0.98	0	0.07	0.01	3.54	0.24	0.82	0.49
	3	0	0.9	0	0.07	0.01	3.85	0.28	0.58	0.42	0.97	0	0.07	0.01	3.94	0.28	0.6	0.46	0.97	0	0.07	0.01	3.95	0.28	0.6	0.45
		1	0.92	0	0.07	0.01	3.74	0.29	0.65	0.45	0.98	0	0.07	0.01	3.73	0.28	0.62	0.44	0.98	0	0.07	0.01	3.73	0.28	0.62	0.44
		2	0.93	0	0.07	0.01	3.71	0.22	0.61	0.38	0.98	0	0.07	0.01	3.7	0.23	0.68	0.39	**0.98**	**0**	**0.07**	**0.01**	**3.7**	**0.23**	**0.7**	**0.38**
		3	0.93	0	0.07	0.01	3.66	0.27	0.73	0.61	0.98	0	0.07	0.01	3.64	0.28	0.73	0.6	0.98	0	0.07	0.01	3.64	0.27	0.73	0.59
		4	0.93	0	0.07	0	3.73	0.19	0.96	0.55	0.98	0	0.07	0	3.7	0.2	0.94	0.62	0.98	0	0.07	0	3.7	0.2	0.94	0.61
	4	0	0.89	0	0.08	0.01	4.12	0.21	0.71	0.49	0.97	0	0.08	0.01	4.17	0.21	0.74	0.48	0.97	0	0.08	0.01	4.16	0.21	0.75	0.48
		1	0.91	0	0.07	0	3.92	0.21	0.78	0.63	0.98	0	0.07	0	3.94	0.2	0.81	0.62	**0.98**	**0**	**0.07**	**0**	**3.94**	**0.2**	**0.92**	**0**
		2	0.07	0.01	3.91	0.26	0.76	0.58	0.98	0	0.07	0.01	3.88	0.28	0.78	0.56	0.98	0	0.07	0.01	3.88	0.28	0.78	0.57	0.92	0
		3	0.07	0.01	3.86	0.24	0.76	0.54	0.98	0	0.07	0.01	3.83	0.23	0.77	0.54	0.98	0	0.07	0.01	3.82	0.23	0.77	0.54	0.93	0
		4	0.07	0.01	3.91	0.2	0.8	0.62	0.98	0	0.07	0.01	3.85	0.19	0.83	0.55	0.98	0	0.07	0.01	3.84	0.2	0.83	0.55	0.89	0
	5	0	0.08	0.01	4.16	0.26	0.65	0.5	0.97	0	0.08	0.01	4.25	0.29	0.7	0.48	0.97	0	0.08	0.01	4.25	0.29	0.71	0.49	0.82	0.63
		1	0.9	0	0.08	0.01	4.04	0.25	0.81	0.54	0.97	0	0.08	0.01	4.05	0.26	0.8	0.52	0.97	0	0.08	0	4.06	0.26	0.8	0.51
		2	0.92	0	0.08	0.01	4.06	0.27	0.63	0.51	0.98	0	0.08	0.01	4.02	0.26	0.67	0.56	0.98	0	0.08	0.01	4.02	0.26	0.67	0.55
		3	0.92	0	0.08	0.01	3.9	0.25	0.91	0.56	0.98	0	0.07	0.01	3.85	0.24	0.86	0.59	0.98	0	0.07	0.01	3.86	0.24	0.87	0.58
		4	0.92	0	0.08	0.01	4.07	0.27	0.65	0.55	**0.98**	**0**	**0.08**	**0.01**	**4**	**0.26**	**0.63**	**0.55**	0.98	0	0.08	0.01	4	0.27	0.63	0.55

Las predicciones para la ciudad de Cuenca teniendo como características principales la profundidad de los árboles de decisión se obtuvieron los resultados mostrados en la tabla 18 donde se resaltan en negrita los mejores resultados para cada día. Para pronosticar el día siguiente se obtuvo mejores resultados con una profundidad de 15 y el retraso de 2 con un R^2=0,98, MAPE=0,06, RMSE=3,3 y PBIAS=0,43, para dos días después, de la misma manera la profundidad fue de 15 con la diferencias en el retraso ya que para este caso fue mejor optar por el retraso de 3 con un R^2=0,98, MAPE=0,07, RMSE=3,43 y PBIAS=0,72, para tres días después se obtuvieron como profundidad 25 y retraso 2 con un R^2=0,98, MAPE=0,07, RMSE=3,7 y PBIAS=0,7, para cuatros días hacia adelante la mejor profundidad fue de 25 y el retraso de 1 con un R^2=0,98, MAPE=0,07, RMSE=3,94 y PBIAS=0,92, finalmente para pronosticar 5 días hacia adelante la mejor estimación fue con una profundidad de 15 tomando el retraso de 4 con un R^2=0,98, MAPE=0,08, RMSE=4 y

PBIAS=0,55. Cabe recalcar que no existe mucha diferencia entre las diferentes estructuras y que mientras aumentan los días a pronosticar la estructura de mayor profundidad se vuelve la mejor opción. Además que mientras más alejado es el día a predecir el error de ajuste tiende a disminuir.

Tabla 19. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para SO_2 de Guayaquil.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.92	0	0.04	0	7.93	0.63	0.41	0.26	0.98	0	0.04	0	8.06	0.65	0.45	0.24	0.98	0	0.04	0	8.05	0.65	0.45	0.24
		1	0.94	0	0.04	0	7.61	0.53	0.44	0.34	0.98	0	0.04	0	7.8	0.5	0.45	0.38	0.98	0	0.04	0	7.8	0.49	0.46	0.37
		2	0.94	0	0.04	0	7.48	0.7	0.53	0.32	0.98	0	0.04	0	7.61	0.69	0.55	0.32	0.98	0	0.04	0	7.6	0.68	0.54	0.32
		3	0.95	0	0.04	0	7.11	0.67	0.43	0.32	**0.98**	**0**	**0.04**	**0**	**7.2**	**0.66**	**0.43**	**0.35**	0.98	0	0.04	0	7.19	0.66	0.43	0.34
		4	0.95	0	0.04	0	7.21	0.65	0.55	0.32	0.98	0	0.04	0	7.26	0.65	0.53	0.33	0.98	0	0.04	0	7.27	0.65	0.53	0.32
	2	0	0.9	0	0.05	0	9.26	0.53	0.55	0.38	0.97	0	0.05	0	9.47	0.56	0.54	0.4	0.97	0	0.05	0	9.47	0.56	0.55	0.41
		1	0.91	0	0.05	0	8.94	0.61	0.59	0.47	0.97	0	0.05	0	9.13	0.61	0.57	0.49	0.97	0	0.05	0	9.13	0.61	0.58	0.49
		2	0.92	0	0.05	0	8.69	0.59	0.6	0.44	0.98	0	0.05	0	8.78	0.57	0.61	0.44	0.98	0	0.05	0	8.78	0.57	0.62	0.45
		3	0.93	0	0.05	0	8.25	0.55	0.56	0.44	**0.98**	**0**	**0.05**	**0**	**8.28**	**0.55**	**0.54**	**0.45**	0.98	0	0.05	0	8.27	0.54	0.55	0.46
		4	0.94	0	0.05	0	8.56	0.52	0.6	0.47	0.98	0	0.05	0	8.53	0.51	0.58	0.45	0.98	0	0.05	0	8.52	0.51	0.58	0.45
	3	0	0.88	0	0.05	0	9.86	0.65	0.5	0.35	0.97	0	0.06	0	10.1	0.68	0.52	0.37	0.97	0	0.06	0	10.1	0.68	0.51	0.37
		1	0.9	0	0.05	0	9.82	0.57	0.72	0.54	0.97	0	0.06	0	9.9	0.54	0.7	0.54	0.97	0	0.06	0	9.89	0.55	0.7	0.55
		2	0.91	0	0.05	0	9.28	0.68	0.59	0.48	**0.97**	**0**	**0.05**	**0**	**9.37**	**0.68**	**0.59**	**0.51**	0.97	0	0.05	0	9.36	0.68	0.6	0.51
		3	0.92	0	0.05	0	9.22	0.6	0.73	0.47	0.97	0	0.05	0	9.24	0.62	0.71	0.48	0.97	0	0.05	0	9.25	0.61	0.71	0.48
		4	0.92	0	0.05	0	9.13	0.66	0.81	0.53	0.97	0	0.05	0	9.08	0.66	0.78	0.52	0.97	0	0.05	0	9.09	0.66	0.79	0.52
	4	0	0.86	0.01	0.06	0	10.8	0.66	0.56	0.52	0.96	0	0.06	0	10.9	0.7	0.56	0.52	0.96	0	0.06	0	10.9	0.7	0.56	0.52
		1	0.88	0	0.06	0	10.6	0.56	0.68	0.45	0.97	0	0.06	0	10.6	0.61	0.69	0.45	0.97	0	0.06	0	10.6	0.62	0.69	0.45
		2	0.89	0.01	0.06	0	10.2	0.65	0.64	0.64	0.97	0	0.06	0	10.2	0.68	0.64	0.66	0.97	0	0.06	0	10.2	0.68	0.64	0.66
		3	0.9	0	0.06	0	10	0.72	0.55	0.39	**0.97**	**0**	**0.06**	**0**	**10**	**0.74**	**0.55**	**0.4**	0.97	0	0.06	0	10	0.73	0.55	0.41
		4	0.91	0	0.06	0	10	0.78	0.67	0.45	0.97	0	0.06	0	9.97	0.77	0.65	0.46	0.97	0	0.06	0	9.97	0.76	0.64	0.45
	5	0	0.84	0.01	0.06	0	11.4	0.75	0.63	0.48	0.96	0	0.07	0	11.5	0.74	0.61	0.47	0.96	0	0.07	0	11.5	0.74	0.61	0.47
		1	0.86	0.01	0.06	0.01	11.3	0.89	0.58	0.5	0.96	0	0.06	0.01	11.4	0.91	0.54	0.5	0.96	0	0.06	0.01	11.4	0.92	0.55	0.5
		2	0.88	0.01	0.06	0.01	11.1	0.88	0.69	0.5	0.96	0	0.06	0.01	11.1	0.89	0.69	0.48	0.96	0	0.06	0.01	11.1	0.88	0.68	0.49
		3	0.89	0.01	0.06	0	10.8	0.86	0.77	0.52	0.96	0	0.06	0.01	10.8	0.86	0.78	0.54	0.96	0	0.06	0	10.8	0.85	0.77	0.54
		4	0.9	0	0.06	0	10.6	0.73	0.72	0.48	**0.96**	**0**	**0.06**	**0**	**10.5**	**0.72**	**0.69**	**0.49**	0.96	0	0.06	0	10.5	0.72	0.69	0.5

Para la ciudad de Guayaquil podemos observar que la mejor estructura fue suficiente la profundidad de 16 para mostrar resultados notables, cuando el número de la profundidad aumenta los modelos no presentan una mejora significativa. Sin embargo, existe diferencia entre el número de retrasos para cada día donde muestra que conforme aumenta el número de días el valor del retraso también aumenta, tal es el caso que para el día 1 con un R^2=0,98, MAPE=0,04, RMSE=7,2 y PBIAS=0,43 donde los dos últimos valores presenta más variaciones, el día 2 el R^2=0,98, MAPE=0,05, RMSE=8,28 y PBIAS=0,54, para el día 3 con un R^2=0,97, MAPE=0,05, RMSE=9,37 y PBIAS=0,59, para el día 4 con un R^2=0,97, MAPE=0,06, RMSE=10 y PBIAS=0,55 y para el día 5 con un R^2=0,96, MAPE=0,06, RMSE=10,5 y PBIAS=0,49. En este caso se puede observar similitud en la estructura para los 5 días de pronóstico.

Tabla 20. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para SO_2 de Quito.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.91	0	0.05	0	3.77	0.22	0.32	0.31	0.98	0	0.05	0	3.84	0.22	0.36	0.29	0.98	0	0.05	0	3.84	0.22	0.36	0.3
		1	0.93	0	0.05	0	3.7	0.14	0.44	0.33	0.98	0	0.05	0	3.73	0.15	0.46	0.31	0.98	0	0.05	0	3.73	0.15	0.46	0.32
		2	0.93	0	0.05	0	3.66	0.23	0.57	0.46	0.98	0	0.05	0	3.68	0.23	0.55	0.44	0.98	0	0.05	0	3.68	0.23	0.55	0.43
		3	0.94	0	0.05	0	3.54	0.25	0.61	0.37	0.98	0	0.05	0	3.56	0.24	0.58	0.37	**0.98**	**0**	**0.05**	**0**	**3.55**	**0.24**	**0.57**	**0.37**
		4	0.95	0	0.05	0	3.51	0.21	0.82	0.36	0.98	0	0.05	0	3.51	0.21	0.82	0.33	0.98	0	0.05	0	3.51	0.21	0.82	0.33
	2	0	0.9	0	0.06	0	4	0.2	0.5	0.34	0.97	0	0.06	0	4.04	0.19	0.5	0.32	0.97	0	0.06	0	4.04	0.2	0.51	0.31
		1	0.92	0	0.05	0	3.84	0.27	0.62	0.43	0.98	0	0.05	0	3.84	0.27	0.62	0.43	0.98	0	0.05	0	3.84	0.27	0.62	0.42
		2	0.93	0	0.05	0	3.84	0.2	0.54	0.45	0.98	0	0.05	0	3.82	0.19	0.55	0.46	0.98	0	0.05	0	3.82	0.2	0.55	0.46
		3	0.93	0	0.05	0	3.83	0.21	0.67	0.44	0.98	0	0.05	0	3.78	0.22	0.64	0.42	**0.98**	**0**	**0.05**	**0**	**3.78**	**0.22**	**0.65**	**0.42**
		4	0.94	0	0.05	0	3.83	0.22	0.56	0.49	0.98	0	0.05	0	3.79	0.2	0.53	0.45	0.98	0	0.05	0	3.79	0.2	0.53	0.45
	3	0	0.9	0	0.06	0	4.09	0.27	0.5	0.35	0.97	0	0.06	0	4.09	0.26	0.5	0.33	0.97	0	0.06	0	4.1	0.26	0.51	0.32
		1	0.91	0	0.06	0	4	0.21	0.55	0.42	0.97	0	0.06	0	4	0.2	0.56	0.43	0.97	0	0.06	0	4	0.2	0.56	0.42
		2	0.92	0	0.06	0	4.03	0.27	0.72	0.52	0.98	0	0.06	0	3.97	0.28	0.7	0.53	0.98	0	0.06	0	3.97	0.28	0.7	0.53
		3	0.93	0	0.06	0	4.08	0.3	0.78	0.66	0.98	0	0.05	0	3.99	0.3	0.76	0.61	**0.98**	**0**	**0.05**	**0**	**3.99**	**0.29**	**0.77**	**0.62**
		4	0.93	0	0.06	0.01	4.07	0.34	0.8	0.64	0.97	0	0.06	0.01	3.99	0.33	0.8	0.65	0.97	0	0.06	0.01	4	0.34	0.81	0.66
	4	0	0.88	0	0.06	0	4.37	0.27	0.62	0.54	0.97	0	0.06	0	4.4	0.26	0.63	0.55	0.97	0	0.06	0	4.4	0.26	0.63	0.55
		1	0.9	0	0.06	0	4.28	0.29	0.58	0.45	0.97	0	0.06	0	4.27	0.29	0.59	0.48	0.97	0	0.06	0	4.27	0.29	0.59	0.48
		2	0.91	0	0.06	0.01	4.33	0.36	0.66	0.49	0.97	0	0.06	0.01	4.27	0.35	0.66	0.46	0.97	0	0.06	0.01	4.28	0.35	0.67	0.48
		3	0.92	0	0.06	0.01	4.32	0.29	0.7	0.58	0.97	0	0.06	0.01	4.26	0.28	0.72	0.58	0.97	0	0.06	0.01	4.26	0.28	0.72	0.58
		4	0.92	0	0.06	0.01	4.19	0.28	0.6	0.49	0.97	0	0.06	0.01	4.12	0.29	0.58	0.47	**0.97**	**0**	**0.06**	**0.01**	**4.12**	**0.29**	**0.59**	**0.46**
	5	0	0.87	0	0.06	0	4.62	0.33	0.61	0.36	0.97	0	0.06	0	4.62	0.33	0.61	0.39	0.97	0	0.06	0	4.62	0.32	0.6	0.39
		1	0.89	0.01	0.06	0	4.44	0.3	0.54	0.4	0.97	0	0.06	0	4.4	0.29	0.57	0.41	0.97	0	0.06	0	4.4	0.29	0.57	0.41
		2	0.9	0	0.06	0.01	4.54	0.32	0.74	0.54	0.97	0	0.06	0.01	4.49	0.32	0.68	0.53	0.97	0	0.06	0.01	4.49	0.31	0.7	0.54
		3	0.91	0	0.06	0	4.44	0.23	0.76	0.47	0.97	0	0.06	0	4.36	0.24	0.71	0.47	**0.97**	**0**	**0.06**	**0**	**4.37**	**0.24**	**0.71**	**0.46**
		4	0.92	0	0.06	0.01	4.38	0.38	0.58	0.51	0.97	0	0.06	0.01	4.32	0.37	0.57	0.51	0.97	0	0.06	0.01	4.32	0.37	0.59	0.51

Sabiendo que la ciudad de Quito se encuentra al norte del Ecuador y es el área más alejada en comparación con las otras dos áreas se pueden observar los siguientes resultados, con poca diferencia entre las diferentes estructuras y tomando en cuenta el PBIAS como determinante se puede decir que la mejor profundidad fue 25 y el mejor retraso para cada día de pronóstico fue de 3. Aquí se puede observar que existe una ligera mejora cuando aumenta el número de neuronas. y que el número de retrasos se mantiene el mismo conforme avanzan los días de pronóstico. Así para el día 1 las métricas de error fueron R^2=0,98, MAPE=0,05, RMSE=3,55 y PBIAS=0,57, para el día dos fueron R^2=0,98, MAPE=0,05, RMSE=3,78 y PBIAS=0,65, donde se puede ver que el PBIAS es el más alto pero el RMSE es más bajo; para el día 3 tenemos R^2=0,98, MAPE=0,05, RMSE=3,99 y PBIAS=0,77; para el cuarto día se puede decir que R^2=0,97, MAPE=0,06, RMSE=4,12 y PBIAS=0,59 donde todas las métricas son las mejores; finalmente para el día 5 podemos observar que el R^2=0,97, MAPE=0,06, RMSE=4,37 y PBIAS=0,71.

6.5.1.2. Predicción de NO_2 mediante Random Forest considerando la profundidad máxima de cada árbol de decisión.

Tabla 21. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para NO_2 de Cuenca.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.88	0	0.08	0.01	0.5	0.03	1.01	0.63	0.97	0	0.09	0.01	0.51	0.03	1.05	0.66	0.97	0	0.09	0.01	0.51	0.03	1.05	0.65
		1	0.89	0	0.08	0.01	0.5	0.03	0.77	0.64	**0.97**	**0**	**0.08**	**0.01**	**0.5**	**0.03**	**0.75**	**0.65**	0.97	0	0.08	0.01	0.5	0.03	0.75	0.66
		2	0.9	0	0.09	0.01	0.49	0.03	0.99	0.77	0.97	0	0.09	0.01	0.5	0.04	0.98	0.76	0.97	0	0.09	0.01	0.5	0.03	0.99	0.76
		3	0.91	0	0.09	0.01	0.5	0.04	0.86	0.58	0.97	0	0.09	0.01	0.5	0.04	0.89	0.58	0.97	0	0.09	0.01	0.5	0.04	0.88	0.58
		4	0.91	0	0.09	0.01	0.5	0.04	0.94	0.71	0.97	0	0.09	0.01	0.5	0.04	0.99	0.75	0.97	0	0.09	0.01	0.5	0.04	0.99	0.74
	2	0	0.86	0.01	0.09	0.01	0.54	0.03	1.09	0.77	0.96	0	0.1	0.01	0.55	0.03	1.08	0.81	0.96	0	0.1	0.01	0.55	0.03	1.08	0.8
		1	0.88	0.01	0.09	0.01	0.53	0.04	0.76	0.63	0.96	0	0.09	0.01	0.54	0.04	0.78	0.59	0.97	0	0.09	0.01	0.54	0.04	0.78	0.59
		2	0.89	0	0.09	0.01	0.53	0.03	0.82	0.8	0.96	0	0.09	0.01	0.53	0.03	0.81	0.79	0.96	0	0.09	0.01	0.53	0.03	0.82	0.78
		3	0.89	0.01	0.09	0.01	0.52	0.05	1	0.76	**0.96**	**0**	**0.09**	**0.01**	**0.52**	**0.05**	**1**	**0.79**	0.96	0	0.09	0.01	0.52	0.05	1	0.79
		4	0.9	0.01	0.09	0.01	0.53	0.03	1.04	0.83	0.96	0	0.09	0.01	0.53	0.03	1.03	0.86	0.96	0	0.09	0.01	0.53	0.03	1.03	0.87
	3	0	0.84	0.01	0.09	0.01	0.56	0.04	0.72	0.51	0.96	0	0.1	0.01	0.57	0.04	0.7	0.56	0.96	0	0.1	0.01	0.57	0.04	0.7	0.56
		1	0.86	0.01	0.1	0.01	0.57	0.04	0.78	0.64	0.96	0	0.1	0.01	0.58	0.03	0.81	0.63	0.96	0	0.1	0.01	0.58	0.03	0.8	0.62
		2	0.88	0.01	0.1	0.01	0.56	0.04	1.04	0.85	0.96	0	0.1	0.01	0.57	0.04	1.01	0.86	0.96	0	0.1	0.01	0.57	0.04	1.01	0.86
		3	0.88	0	0.1	0.01	0.55	0.03	1.06	0.91	**0.96**	**0**	**0.1**	**0.01**	**0.55**	**0.03**	**1.09**	**0.89**	0.96	0	0.1	0.01	0.56	0.03	1.09	0.9
		4	0.89	0.01	0.1	0.01	0.55	0.04	1.21	0.98	0.96	0	0.1	0.01	0.56	0.04	1.21	0.97	0.96	0	0.1	0.01	0.56	0.04	1.22	0.97
	4	0	0.82	0.01	0.1	0.01	0.6	0.04	0.97	0.74	0.95	0	0.11	0.01	0.62	0.04	0.95	0.73	0.96	0	0.11	0.01	0.62	0.04	0.96	0.75
		1	0.85	0.01	0.1	0.01	0.59	0.04	1.11	0.92	0.96	0	0.1	0.01	0.59	0.04	1.12	0.87	0.96	0	0.1	0.01	0.59	0.04	1.1	0.87
		2	0.87	0	0.1	0.01	0.58	0.03	1.19	0.88	0.96	0	0.1	0.01	0.58	0.03	1.19	0.9	0.96	0	0.1	0.01	0.58	0.03	1.18	0.88
		3	0.88	0.01	0.1	0.01	0.56	0.04	0.92	0.57	0.96	0	0.1	0.01	0.56	0.04	0.87	0.55	**0.96**	**0**	**0.1**	**0.01**	**0.56**	**0.04**	**0.87**	**0.54**
		4	0.89	0.01	0.1	0.01	0.55	0.04	1.07	0.77	0.96	0	0.1	0.01	0.55	0.04	1.05	0.76	0.96	0	0.1	0.01	0.55	0.05	1.08	0.75
	5	0	0.82	0.01	0.1	0.01	0.6	0.03	0.85	0.71	0.95	0	0.1	0.01	0.61	0.03	0.89	0.69	0.95	0	0.1	0.01	0.61	0.03	0.9	0.69
		1	0.85	0.01	0.1	0.01	0.57	0.03	1.11	0.63	0.96	0	0.1	0.01	0.57	0.02	1.14	0.66	0.96	0	0.1	0.01	0.57	0.02	1.14	0.67
		2	0.86	0	0.1	0.01	0.58	0.03	1.3	1.07	0.96	0	0.1	0.01	0.57	0.03	1.27	1.06	0.96	0	0.1	0.01	0.57	0.03	1.27	1.06
		3	0.88	0.01	0.1	0.01	0.56	0.04	0.81	0.62	**0.96**	**0**	**0.1**	**0.01**	**0.55**	**0.03**	**0.74**	**0.62**	0.96	0	0.1	0.01	0.55	0.03	0.75	0.65
		4	0.88	0.01	0.1	0.01	0.54	0.04	1.18	0.86	0.96	0	0.09	0.01	0.54	0.04	1.13	0.85	0.96	0	0.09	0.01	0.54	0.04	1.12	0.85

En el análisis del gas contaminante NO_2 podemos expresar lo siguiente: para el día 1 la profundidad adecuada fue de 15 con un retraso óptimo de 1 con una R^2=0,97, MAPE=0,08, RMSE=0,5 y PBIAS=0,75, para el día 2 de la misma manera la profundidad fue de 15 con el retraso de 3 con una R^2=0,96, MAPE=0,09, RMSE=0,52 y PBIAS=1; para el día 4 el valor de la profundidad fue de 25 con un retraso de 3 con una R^2=0,96, MAPE=0,1, RMSE=0,55 y PBIAS=1,09, aquí se tomó en cuenta la desviación estándar como determinante ya que los valores no varían mucho. Finalmente para el día 5 la profundidad óptima fue de 15 con el retraso de 3 con un R^2=0,96, MAPE=0,1, RMSE=0,55 y PBIAS=0,74 donde todas las métricas fueron muy buenas.

Tabla 22. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para NO_2 de Guayaquil.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.88	0.01	0.06	0	0.68	0.05	0.62	0.46	0.97	0	0.06	0	0.69	0.05	0.61	0.45	0.97	0	0.06	0	0.69	0.05	0.61	0.44
		1	0.91	0	0.06	0	0.63	0.04	0.74	0.66	0.97	0	0.06	0	0.64	0.04	0.77	0.69	0.97	0	0.06	0	0.63	0.04	0.76	0.67
		2	0.92	0	0.06	0	0.63	0.04	0.8	0.59	0.98	0	0.06	0	0.64	0.04	0.79	0.61	0.98	0	0.06	0	0.64	0.04	0.78	0.61
		3	0.94	0	0.06	0	0.59	0.05	0.61	0.42	**0.98**	**0**	**0.06**	**0**	**0.59**	**0.05**	**0.62**	**0.41**	0.98	0	0.06	0	0.59	0.05	0.62	0.41
		4	0.94	0	0.06	0	0.58	0.04	0.77	0.51	0.98	0	0.06	0	0.59	0.04	0.75	0.52	0.98	0	0.06	0	0.59	0.04	0.76	0.51
	2	0	0.86	0.01	0.07	0	0.75	0.06	0.81	0.6	0.96	0	0.07	0	0.76	0.06	0.81	0.58	0.96	0	0.07	0	0.76	0.06	0.8	0.57
		1	0.89	0	0.07	0	0.72	0.05	0.77	0.63	0.97	0	0.07	0	0.72	0.05	0.78	0.62	0.97	0	0.07	0	0.72	0.05	0.79	0.63
		2	0.91	0	0.06	0	0.68	0.05	0.81	0.66	0.97	0	0.06	0	0.68	0.05	0.83	0.68	0.97	0	0.06	0	0.68	0.05	0.82	0.67
		3	0.91	0.01	0.06	0	0.69	0.05	0.6	0.52	**0.97**	**0**	**0.06**	**0.01**	**0.69**	**0.05**	**0.59**	**0.55**	0.97	0	0.06	0	0.69	0.05	0.6	0.55
		4	0.92	0.01	0.06	0.01	0.67	0.06	0.94	0.63	0.97	0	0.06	0.01	0.66	0.06	0.93	0.61	0.97	0	0.06	0.01	0.66	0.06	0.95	0.62
	3	0	0.84	0.01	0.07	0	0.79	0.04	0.94	0.53	0.96	0	0.07	0	0.8	0.04	0.92	0.54	0.96	0	0.07	0	0.8	0.04	0.92	0.53
		1	0.87	0	0.07	0	0.76	0.04	0.82	0.68	0.96	0	0.07	0	0.77	0.04	0.82	0.66	0.96	0	0.07	0	0.77	0.04	0.82	0.67
		2	0.88	0.01	0.07	0	0.74	0.05	0.85	0.55	0.96	0	0.07	0.01	0.75	0.05	0.84	0.53	0.97	0	0.07	0.01	0.75	0.05	0.82	0.53
		3	0.9	0.01	0.07	0	0.74	0.05	0.8	0.62	**0.97**	**0**	**0.07**	**0.01**	**0.74**	**0.05**	**0.79**	**0.58**	0.97	0	0.07	0.01	0.74	0.05	0.79	0.59
		4	0.91	0.01	0.07	0.01	0.74	0.06	1.02	0.87	0.97	0	0.07	0.01	0.73	0.05	1.02	0.87	0.97	0	0.07	0.01	0.73	0.05	1.03	0.87
	4	0	0.81	0.01	0.08	0	0.83	0.05	0.69	0.57	0.95	0	0.08	0	0.85	0.05	0.75	0.56	0.95	0	0.08	0	0.85	0.05	0.75	0.57
		1	0.84	0.01	0.08	0	0.83	0.05	0.81	0.65	0.96	0	0.08	0	0.83	0.05	0.81	0.64	0.96	0	0.08	0	0.83	0.05	0.81	0.63
		2	0.87	0.01	0.08	0	0.8	0.05	0.95	0.64	0.96	0	0.08	0	0.79	0.04	0.94	0.6	0.96	0	0.08	0	0.79	0.04	0.94	0.6
		3	0.89	0.01	0.08	0.01	0.79	0.05	0.81	0.68	**0.96**	**0**	**0.07**	**0.01**	**0.78**	**0.05**	**0.82**	**0.62**	0.96	0	0.07	0.01	0.78	0.05	0.82	0.63
		4	0.89	0.01	0.07	0.01	0.76	0.05	0.92	0.64	0.96	0	0.07	0.01	0.75	0.05	0.98	0.66	0.96	0	0.07	0.01	0.75	0.05	0.96	0.66
	5	0	0.8	0.01	0.08	0.01	0.86	0.05	0.72	0.56	0.95	0	0.08	0.01	0.87	0.05	0.75	0.57	0.95	0	0.08	0.01	0.87	0.05	0.74	0.57
		1	0.84	0.01	0.08	0	0.83	0.04	0.8	0.79	0.96	0	0.08	0	0.83	0.04	0.82	0.82	0.96	0	0.08	0	0.83	0.04	0.82	0.82
		2	0.87	0.01	0.08	0.01	0.82	0.05	0.81	0.58	0.96	0	0.08	0.01	0.82	0.05	0.81	0.59	0.96	0	0.08	0.01	0.82	0.05	0.81	0.59
		3	0.88	0.01	0.08	0.01	0.8	0.05	0.81	0.64	0.96	0	0.08	0.01	0.79	0.05	0.86	0.63	0.96	0	0.08	0.01	0.8	0.05	0.86	0.62
		4	0.89	0.01	0.07	0	0.77	0.05	0.89	0.66	**0.96**	**0**	**0.07**	**0**	**0.76**	**0.05**	**0.87**	**0.65**	0.96	0	0.07	0	0.76	0.05	0.87	0.64

En el caso de Guayaquil, la profundidad determinante fue de 15 mostrando mejores resultados, además teniendo en cuenta el número de retrasos también se puede apreciar que para los días 1, 2 , 3, 4 el retraso fue de 3 con una R^2=0,98, MAPE=0,06, RMSE=0,59 y PBIAS=0,62; R^2=0,97, MAPE=0,06, RMSE=0,69 y PBIAS=0,59; R^2=0,97, MAPE=0,07, RMSE=0,74 y PBIAS=0,79; y R^2=0,96, MAPE=0,07, RMSE=0,78 y PBIAS=0,82 respectivamente, con la diferencia para el día 5 donde el retraso aumentó a 4 con un R^2=0,96, MAPE=0,07, RMSE=0,77 y PBIAS=0,87. Podemos observar que el número de retraso para este caso se mantiene constante los primeros días de pronóstico y el último día el retraso aumenta.

Tabla 23. Resultados de análisis de Random Forest tomando en cuenta la profundidad máxima para NO_2 de Quito.

			Profundidad de 5								Profundidad de 15								Profundidad de 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.86	0.01	0.07	0	0.79	0.05	0.77	0.57	0.96	0	0.07	0	0.81	0.05	0.82	0.58	0.97	0	0.07	0	0.81	0.05	0.83	0.58
		1	0.88	0.01	0.07	0.01	0.77	0.05	0.73	0.61	**0.97**	**0**	**0.07**	**0.01**	**0.78**	**0.05**	**0.8**	**0.63**	0.97	0	0.07	0.01	0.78	0.06	0.8	0.63
		2	0.9	0.01	0.07	0.01	0.8	0.05	0.75	0.57	0.97	0	0.07	0.01	0.81	0.05	0.81	0.57	0.97	0	0.07	0.01	0.81	0.05	0.8	0.58
		3	0.9	0.01	0.07	0.01	0.77	0.06	0.96	0.67	0.97	0	0.07	0.01	0.78	0.06	0.97	0.66	0.97	0	0.07	0.01	0.78	0.06	0.98	0.65
		4	0.91	0.01	0.08	0.01	0.81	0.07	0.98	0.71	0.97	0	0.08	0.01	0.82	0.07	1	0.72	0.97	0	0.08	0.01	0.82	0.07	0.99	0.72
	2	0	0.82	0.01	0.08	0.01	0.92	0.06	0.85	0.66	0.95	0	0.08	0.01	0.93	0.06	0.87	0.7	0.95	0	0.08	0.01	0.93	0.06	0.88	0.69
		1	0.84	0.01	0.08	0.01	0.9	0.06	0.69	0.64	0.95	0	0.08	0.01	0.91	0.06	0.7	0.69	0.96	0	0.08	0.01	0.91	0.06	0.7	0.66
		2	0.86	0.01	0.08	0.01	0.91	0.07	1.13	0.74	0.96	0	0.08	0.01	0.91	0.07	1.13	0.76	0.96	0	0.08	0.01	0.91	0.07	1.12	0.76
		3	0.87	0.01	0.08	0.01	0.89	0.06	1.01	0.73	0.96	0	0.08	0.01	0.89	0.06	1.05	0.74	0.96	0	0.08	0.01	0.89	0.06	1.03	0.75
		4	0.88	0.01	0.08	0.01	0.87	0.07	0.85	0.65	**0.96**	**0**	**0.08**	**0.01**	**0.86**	**0.07**	**0.84**	**0.57**	0.96	0	0.08	0.01	0.86	0.07	0.83	0.58
	3	0	0.79	0.01	0.09	0.01	1	0.06	1.04	0.73	0.95	0	0.09	0.01	1.02	0.06	1.09	0.81	0.95	0	0.09	0.01	1.02	0.06	1.08	0.82
		1	0.81	0.01	0.09	0.01	0.96	0.05	0.93	0.67	0.95	0	0.09	0.01	0.96	0.05	0.93	0.67	0.95	0	0.09	0.01	0.96	0.05	0.94	0.68
		2	0.83	0.01	0.09	0.01	0.95	0.06	1.19	0.66	0.95	0	0.09	0.01	0.95	0.07	1.17	0.66	0.95	0	0.09	0.01	0.95	0.07	1.17	0.68
		3	0.85	0.01	0.09	0.01	0.93	0.06	1.09	0.76	0.95	0	0.09	0.01	0.92	0.06	1.07	0.77	0.95	0	0.09	0.01	0.92	0.06	1.05	0.77
		4	0.87	0.01	0.09	0.01	0.93	0.06	1.06	0.63	**0.96**	**0**	**0.09**	**0.01**	**0.92**	**0.06**	**0.98**	**0.61**	0.96	0	0.09	0.01	0.92	0.06	0.98	0.61
	4	0	0.78	0.01	0.09	0.01	0.99	0.05	0.94	0.69	0.94	0	0.09	0.01	1.02	0.05	0.89	0.67	0.94	0	0.09	0.01	1.02	0.05	0.89	0.67
		1	0.81	0.01	0.09	0.01	0.97	0.07	1.11	0.76	0.95	0	0.09	0.01	0.98	0.07	1.08	0.75	0.95	0	0.09	0.01	0.98	0.07	1.09	0.77
		2	0.84	0.01	0.09	0.01	0.97	0.08	1.23	0.91	0.95	0	0.09	0.01	0.97	0.08	1.21	0.89	0.95	0	0.09	0.01	0.97	0.08	1.22	0.9
		3	0.85	0.01	0.09	0.01	0.95	0.05	0.97	0.77	0.95	0	0.09	0.01	0.95	0.05	0.96	0.75	0.95	0	0.09	0.01	0.95	0.05	0.97	0.76
		4	0.87	0.01	0.09	0.01	0.94	0.07	1.03	0.73	**0.96**	**0**	**0.09**	**0.01**	**0.93**	**0.07**	**0.96**	**0.74**	0.96	0	0.09	0.01	0.93	0.07	0.95	0.74
	5	0	0.79	0.01	0.09	0.01	0.98	0.07	0.64	0.61	0.95	0	0.09	0.01	0.99	0.06	0.66	0.6	0.95	0	0.09	0.01	0.99	0.06	0.66	0.61
		1	0.82	0.01	0.09	0.01	0.95	0.06	0.88	0.63	0.95	0	0.09	0.01	0.95	0.06	0.88	0.63	0.95	0	0.09	0.01	0.95	0.06	0.87	0.62
		2	0.84	0.01	0.09	0.01	0.96	0.06	1.04	0.79	0.95	0	0.09	0.01	0.95	0.06	1.04	0.78	0.95	0	0.09	0.01	0.95	0.06	1.03	0.78
		3	0.86	0.01	0.09	0.01	0.93	0.06	0.95	0.65	0.95	0	0.09	0.01	0.93	0.06	0.94	0.63	0.95	0	0.09	0.01	0.93	0.06	0.93	0.63
		4	0.87	0.01	0.09	0.01	0.92	0.07	0.94	0.71	**0.96**	**0**	**0.09**	**0.01**	**0.91**	**0.07**	**0.92**	**0.71**	0.96	0	0.09	0.01	0.91	0.07	0.92	0.72

El análisis mostrado en la tabla se puede apreciar que para este caso la profundidad adecuada fue 15 para todos los días a predecir, mostrando una predominancia de retraso en 4 con excepción para el día 1 el cual el retraso mostró un valor de 1 lo que indica que para este conjunto de datos no fue necesario aumentar en gran medida el valor de la profundidad máxima. Así, podemos decir que para el día 1 las métricas fueron: R^2=0,97, MAPE=0,07, RMSE=0,78 y PBIAS=0,8, para el día 2 tenemos un R^2=0,96, MAPE=0,08, RMSE=0,86 y PBIAS=0,84, para el día 3 tenemos un R^2=0,96, MAPE=0,09, RMSE=0,92 y PBIAS=0,98, para el día 4 fue el R^2=0,96, MAPE=0,09, RMSE=0,93 y PBIAS=0,96 y para el días 5 los mejores valores de error fueron R^2=0,96, MAPE=0,09, RMSE=0,91 y PBIAS=0,92. Finalmente podemos darnos cuenta que no existe gran variación entre el valor de la profundidad 15 y 25 lo que se podría esperar que mientras más aumenta la profundidad podría encontrarnos con un sobreajuste del modelo y disminuir las métricas de bondad de ajuste y aumentar el error.

6.5.1.3. Predicción del SO_2 mediante Random Forest considerando el número mínimo de muestras en los nodos hoja.

La segunda característica que se tomó en cuenta fue el número mínimo de muestras para considerarse como nodo hoja donde se determina la extensión de los árboles de decisión. Para ellos tenemos los siguientes resultados:

Tabla 24. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para SO_2 de Cuenca.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.94	0	0.06	0.01	3.49	0.23	0.53	0.39	0.89	0	0.06	0.01	3.52	0.22	0.55	0.35	0.87	0	0.07	0.01	3.64	0.23	0.54	0.32
		1	0.95	0	0.06	0	3.4	0.2	0.63	0.38	0.9	0	0.07	0.01	3.53	0.25	0.61	0.34	0.88	0.01	0.07	0.01	3.67	0.28	0.62	0.4
		2	**0.96**	**0**	**0.06**	**0**	**3.27**	**0.2**	**0.44**	**0.44**	0.91	0	0.06	0	3.41	0.21	0.5	0.47	0.89	0	0.07	0.01	3.54	0.23	0.56	0.48
		3	0.96	0	0.06	0.01	3.19	0.2	0.75	0.43	0.91	0	0.06	0.01	3.35	0.23	0.79	0.48	0.88	0	0.07	0.01	3.48	0.3	0.76	0.62
		4	0.96	0	0.06	0	3.25	0.16	0.51	0.43	0.91	0	0.07	0	3.41	0.2	0.57	0.42	0.88	0	0.07	0.01	3.62	0.25	0.6	0.42
	2	0	0.93	0	0.07	0.01	3.61	0.24	0.57	0.41	0.88	0	0.07	0.01	3.68	0.23	0.59	0.37	0.85	0	0.07	0.01	3.88	0.22	0.63	0.4
		1	0.94	0	0.06	0.01	3.49	0.26	0.71	0.69	0.89	0	0.07	0.01	3.68	0.28	0.73	0.72	0.86	0.01	0.07	0.01	3.85	0.31	0.71	0.68
		2	**0.95**	**0**	**0.07**	**0**	**3.52**	**0.17**	**0.65**	**0.63**	0.9	0	0.07	0	3.69	0.19	0.64	0.67	0.87	0	0.07	0.01	3.81	0.2	0.71	0.66
		3	0.95	0	0.07	0.01	3.43	0.23	0.67	0.53	0.89	0	0.07	0.01	3.66	0.27	0.74	0.49	0.87	0	0.07	0.01	3.77	0.31	0.71	0.48
		4	0.95	0	0.07	0.01	3.53	0.23	0.8	0.5	0.89	0	0.07	0.01	3.75	0.26	0.71	0.46	0.86	0	0.08	0.01	3.91	0.26	0.73	0.51
	3	0	0.92	0	0.07	0.01	3.85	0.28	0.59	0.39	0.86	0.01	0.07	0.01	3.93	0.29	0.61	0.41	0.83	0.01	0.08	0.01	4.1	0.3	0.62	0.44
		1	0.93	0	0.07	0.01	3.73	0.29	0.63	0.45	0.87	0.01	0.07	0.01	3.89	0.27	0.68	0.52	0.84	0.01	0.08	0.01	4.13	0.29	0.67	0.54
		2	**0.95**	**0**	**0.07**	**0.01**	**3.69**	**0.22**	**0.61**	**0.37**	0.88	0.01	0.08	0.01	3.9	0.25	0.54	0.38	0.85	0.01	0.08	0.01	4.13	0.3	0.6	0.44
		3	0.95	0	0.07	0.01	3.64	0.28	0.76	0.59	0.88	0	0.07	0.01	3.88	0.3	0.89	0.61	0.84	0.01	0.08	0.01	4.06	0.28	0.93	0.6
		4	0.95	0	0.07	0	3.72	0.2	0.97	0.6	0.88	0	0.08	0.01	3.97	0.23	0.97	0.58	0.84	0.01	0.08	0.01	4.19	0.25	0.97	0.61
	4	0	0.92	0	0.08	0.01	4.12	0.2	0.69	0.49	0.85	0.01	0.08	0.01	4.18	0.24	0.66	0.49	0.82	0	0.08	0.01	4.38	0.28	0.7	0.57
		1	0.93	0	0.07	0	3.91	0.22	0.79	0.62	0.87	0.01	0.08	0.01	4.13	0.25	0.7	0.63	0.83	0.01	0.08	0.01	4.32	0.26	0.69	0.63
		2	0.94	0	0.07	0.01	3.89	0.28	0.77	0.55	0.87	0.01	0.08	0.01	4.14	0.31	0.74	0.52	0.84	0.01	0.08	0.01	4.28	0.3	0.72	0.53
		3	**0.94**	**0**	**0.07**	**0.01**	**3.85**	**0.22**	**0.75**	**0.56**	0.87	0	0.08	0.01	4.11	0.29	0.79	0.71	0.84	0.01	0.08	0.01	4.2	0.28	0.82	0.71
		4	0.94	0	0.07	0.01	3.88	0.19	0.8	0.56	0.87	0	0.08	0.01	4.22	0.26	0.84	0.6	0.84	0.01	0.08	0.01	4.36	0.29	0.88	0.61
	5	0	0.91	0	0.08	0.01	4.2	0.27	0.69	0.48	0.85	0.01	0.08	0.01	4.25	0.26	0.7	0.44	0.82	0.01	0.08	0.01	4.43	0.26	0.71	0.52
		1	0.93	0	0.08	0.01	4.02	0.25	0.78	0.53	0.86	0.01	0.08	0.01	4.21	0.25	0.79	0.56	0.83	0.01	0.08	0.01	4.34	0.24	0.75	0.59
		2	0.94	0	0.08	0.01	4.02	0.26	0.67	0.54	0.87	0	0.08	0.01	4.23	0.29	0.64	0.56	0.83	0.01	0.08	0.01	4.36	0.3	0.73	0.57
		3	0.94	0	0.07	0.01	3.86	0.25	0.9	0.59	0.87	0	0.08	0.01	4.08	0.28	0.94	0.66	0.83	0.01	0.08	0.01	4.22	0.3	0.96	0.72
		4	**0.94**	**0**	**0.08**	**0.01**	**4.04**	**0.27**	**0.65**	**0.54**	0.86	0.01	0.09	0.01	4.37	0.25	0.74	0.57	0.83	0.01	0.09	0.01	4.52	0.25	0.73	0.58

Para esta característica podemos ver una clara inclinación en los resultados donde los mejores resultados fueron con un mínimo de muestras de 5 para todos los días pronosticados teniendo como retrasos los siguiente: para el día 1, 2 y 3, el retraso fue de 2, con un R^2=0,96, MAPE=0,6, RMSE=3,27 y PBIAS=0,44; R2=0,95, MAPE=0,07, RMSE=3,52 y PBIAS=0,65; R^2=0,95, MAPE=0,07, RMSE=3,69 y PBIAS=0,61 respectivamente, para el días 4 el retraso fue de 3 con un R^2=0,94, MAPE=0,07, RMSE=3,85 y PBIAS=0,75, finalmente para el día 5 el mejor retraso fue de 4 con un R^2=0,94, MAPE=0,08, RMSE=,04 y PBIAS=0,65. Existe una clara diferencia entre las diferentes estructuras y se puede determinar que mientras el número de muestras aumenta las métricas de rendimiento del pronóstico disminuye notablemente y que es un claro ejemplo que el mayor número de numero de muestras no mejora el resultado, por el contrario puede llevar a provocar un sobre ajuste del modelo.

Tabla 25. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para SO_2 de Guayaquil.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.93	0	0.04	0	7.96	0.61	0.46	0.3	0.89	0	0.04	0	8.11	0.66	0.46	0.32	0.87	0	0.04	0	8.17	0.7	0.45	0.31
		1	0.94	0	0.04	0	7.69	0.53	0.43	0.33	0.9	0	0.04	0	7.72	0.63	0.46	0.34	0.89	0	0.04	0	7.81	0.68	0.41	0.36
		2	0.95	0	0.04	0	7.55	0.65	0.52	0.31	0.9	0	0.04	0	7.56	0.68	0.54	0.32	0.89	0	0.04	0	7.62	0.74	0.53	0.34
		3	**0.95**	**0**	**0.04**	**0**	**7.14**	**0.69**	**0.44**	**0.34**	0.91	0	0.04	0	7.12	0.69	0.44	0.33	0.9	0.01	0.04	0	7.19	0.68	0.43	0.34
		4	0.96	0	0.04	0	7.17	0.64	0.53	0.31	0.92	0	0.04	0	7.13	0.67	0.52	0.3	0.9	0	0.04	0	7.23	0.71	0.51	0.35
	2	0	0.91	0	0.05	0	9.35	0.55	0.53	0.39	0.86	0	0.05	0	9.45	0.6	0.57	0.4	0.84	0.01	0.05	0	9.51	0.61	0.57	0.38
		1	0.93	0	0.05	0	9.02	0.63	0.6	0.47	0.87	0	0.05	0	9.07	0.7	0.59	0.48	0.85	0.01	0.05	0	9.1	0.75	0.62	0.49
		2	0.94	0	0.05	0	8.74	0.59	0.59	0.41	0.88	0	0.05	0	8.73	0.66	0.58	0.4	0.86	0.01	0.05	0	8.81	0.65	0.61	0.41
		3	**0.95**	**0**	**0.04**	**0**	**7.14**	**0.69**	**0.44**	**0.34**	0.91	0	0.04	0	7.12	0.69	0.44	0.33	0.9	0.01	0.04	0	7.19	0.68	0.43	0.34
		4	0.94	0	0.05	0	8.54	0.53	0.59	0.47	0.89	0	0.05	0	8.58	0.54	0.63	0.48	0.87	0	0.05	0	8.63	0.54	0.63	0.5
	3	0	0.9	0	0.05	0	9.93	0.65	0.52	0.39	0.84	0	0.06	0	10	0.68	0.46	0.39	0.82	0.01	0.06	0	10.1	0.72	0.44	0.38
		1	0.92	0	0.05	0	9.79	0.54	0.71	0.55	0.85	0	0.06	0	9.82	0.53	0.7	0.57	0.83	0	0.06	0	9.88	0.57	0.71	0.6
		2	**0.93**	**0**	**0.05**	**0**	**9.37**	**0.67**	**0.61**	**0.48**	0.86	0.01	0.05	0	9.4	0.72	0.58	0.48	0.84	0.01	0.05	0	9.46	0.76	0.59	0.46
		3	0.93	0	0.05	0	9.26	0.58	0.74	0.49	0.87	0	0.05	0	9.34	0.56	0.75	0.48	0.84	0	0.05	0	9.42	0.61	0.72	0.47
		4	0.94	0	0.05	0	9.1	0.63	0.79	0.5	0.87	0	0.05	0	9.27	0.64	0.83	0.54	0.84	0.01	0.05	0	9.38	0.65	0.85	0.56
	4	0	0.88	0	0.06	0	10.8	0.68	0.58	0.5	0.8	0.01	0.06	0	10.9	0.62	0.53	0.48	0.77	0.01	0.06	0	11	0.61	0.51	0.48
		1	0.9	0	0.06	0	10.5	0.58	0.71	0.42	0.82	0.01	0.06	0	10.6	0.53	0.67	0.44	0.79	0.01	0.06	0	10.8	0.54	0.66	0.46
		2	0.91	0	0.06	0	10.3	0.69	0.66	0.66	0.83	0.01	0.06	0	10.4	0.69	0.67	0.62	0.8	0.01	0.06	0	10.6	0.73	0.67	0.6
		3	**0.92**	**0**	**0.06**	**0**	**10.1**	**0.67**	**0.57**	**0.4**	0.84	0	0.06	0	10.4	0.65	0.55	0.43	0.8	0.01	0.06	0	10.5	0.63	0.56	0.43
		4	0.92	0	0.06	0	10.1	0.76	0.67	0.44	0.84	0.01	0.06	0	10.4	0.76	0.72	0.51	0.81	0.01	0.06	0	10.6	0.78	0.75	0.54
	5	0	0.87	0	0.07	0	11.4	0.71	0.61	0.47	0.78	0.01	0.07	0	11.6	0.7	0.61	0.47	0.75	0.01	0.07	0	11.7	0.7	0.6	0.44
		1	0.89	0	0.06	0.01	11.4	0.9	0.57	0.52	0.8	0.01	0.06	0.01	11.6	0.92	0.59	0.56	0.77	0.01	0.06	0.01	11.7	0.92	0.62	0.57
		2	0.9	0	0.06	0.01	11.2	0.87	0.72	0.5	0.81	0.01	0.06	0.01	11.4	0.86	0.67	0.5	0.78	0.01	0.06	0.01	11.5	0.88	0.67	0.5
		3	0.91	0	0.06	0.01	10.9	0.86	0.76	0.53	0.82	0.01	0.06	0	11.1	0.8	0.78	0.5	0.78	0.01	0.06	0	11.3	0.78	0.77	0.53
		4	**0.91**	**0**	**0.06**	**0**	**10.6**	**0.72**	**0.72**	**0.46**	0.82	0.01	0.06	0	11	0.66	0.74	0.47	0.78	0.01	0.06	0	11.1	0.61	0.73	0.5

Para esta área de estudio se puede observar que el mejor número mínimo de muestras para considerarse nodo hoja fue 5, diferenciándose por el número de retrasos teniendo lo siguiente: para el día 1 y 2 el mejor retraso fue de 3 con un R^2=0,95, MAPE=0,04, RMSE=7,14 y PBIAS=0,44 y R^2=0,95, MAPE=0,04, RMSE=7,14 y PBIAS=0,44 respectivamente, para el día el días 3 el mejor retraso fue de 2 con un R^2=0,93, MAPE=0,05, RMSE=9,37 y PBIAS=0,61, para el día 4 de la misma manera el retraso fue 3 con un R^2=0,92, MAPE=0,06, RMSE=10,1 y PBIAS=0,57 y para el día 5 el mejor retraso fue de 4 con un R^2=0,91, MAPE=0,06, RMSE=10,6 y PBIAS=0,72. Podemos apreciar que de la misma manera, conforme aumenta el número de muestras, la bondad de ajuste del modelo disminuye y considerando las variables ficticias o número de retrasos el rendimiento disminuye conforme estas aumentan. Además la bondad de ajuste disminuye ligeramente mientras aumenta el día de predicciones.

Tabla 26. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para SO_2 de Quito.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.93	0	0.05	0	3.77	0.22	0.34	0.29	0.88	0	0.05	0	3.82	0.23	0.33	0.3	0.86	0	0.05	0	3.92	0.24	0.37	0.32
		1	0.94	0	0.05	0	3.71	0.16	0.46	0.29	0.89	0	0.05	0	3.8	0.18	0.46	0.32	0.86	0	0.05	0	3.92	0.19	0.47	0.34
		2	0.94	0	0.05	0	3.68	0.24	0.58	0.43	0.89	0	0.05	0	3.74	0.24	0.6	0.48	0.87	0	0.05	0	3.82	0.26	0.6	0.45
		3	**0.95**	**0**	**0.05**	**0**	**3.56**	**0.25**	**0.59**	**0.36**	0.89	0	0.05	0	3.63	0.27	0.64	0.37	0.87	0.01	0.05	0	3.74	0.29	0.62	0.39
		4	0.95	0	0.05	0	3.55	0.2	0.83	0.36	0.89	0	0.05	0	3.62	0.23	0.82	0.39	0.86	0.01	0.05	0	3.75	0.27	0.85	0.44
	2	0	0.92	0	0.06	0	4.03	0.19	0.5	0.29	0.86	0	0.06	0	4.16	0.22	0.58	0.38	0.83	0	0.06	0	4.25	0.25	0.61	0.41
		1	0.93	0	0.05	0	3.86	0.27	0.63	0.44	0.87	0	0.06	0	3.94	0.28	0.61	0.48	0.85	0.01	0.06	0	4.03	0.3	0.63	0.5
		2	**0.94**	**0**	**0.05**	**0**	**3.83**	**0.19**	**0.55**	**0.46**	0.88	0	0.05	0	3.93	0.25	0.6	0.42	0.85	0	0.06	0	4.02	0.27	0.6	0.44
		3	0.94	0	0.05	0	3.81	0.24	0.66	0.44	0.88	0.01	0.05	0	3.96	0.26	0.66	0.41	0.85	0.01	0.06	0	4.06	0.27	0.65	0.37
		4	0.94	0	0.05	0	3.84	0.22	0.59	0.47	0.88	0	0.06	0	3.96	0.25	0.62	0.49	0.84	0.01	0.06	0	4.07	0.27	0.64	0.52
	3	0	0.91	0	0.06	0	4.07	0.26	0.51	0.34	0.85	0.01	0.06	0	4.16	0.31	0.52	0.39	0.83	0.01	0.06	0	4.27	0.34	0.56	0.42
		1	**0.93**	**0**	**0.06**	**0**	**4**	**0.2**	**0.54**	**0.42**	0.87	0.01	0.06	0	4.13	0.22	0.54	0.44	0.84	0.01	0.06	0	4.23	0.24	0.57	0.44
		2	0.94	0	0.06	0	4.01	0.25	0.72	0.56	0.87	0	0.06	0	4.16	0.24	0.8	0.55	0.84	0.01	0.06	0	4.26	0.24	0.82	0.54
		3	0.94	0	0.06	0	4.07	0.27	0.78	0.63	0.87	0	0.06	0	4.2	0.25	0.73	0.67	0.84	0.01	0.06	0.01	4.32	0.28	0.73	0.7
		4	0.94	0	0.06	0.01	4.06	0.35	0.83	0.63	0.87	0.01	0.06	0.01	4.17	0.35	0.81	0.61	0.83	0.01	0.06	0.01	4.32	0.35	0.85	0.6
	4	0	0.9	0	0.06	0	4.39	0.27	0.61	0.53	0.84	0.01	0.06	0	4.5	0.29	0.69	0.53	0.81	0.01	0.07	0.01	4.65	0.31	0.72	0.54
		1	0.92	0	0.06	0.01	4.31	0.29	0.57	0.45	0.85	0.01	0.06	0	4.45	0.3	0.61	0.42	0.82	0.01	0.06	0.01	4.54	0.3	0.64	0.41
		2	0.93	0	0.06	0.01	4.34	0.35	0.65	0.5	0.86	0	0.06	0.01	4.5	0.34	0.72	0.51	0.83	0.01	0.07	0.01	4.59	0.36	0.71	0.52
		3	0.93	0	0.06	0.01	4.32	0.28	0.73	0.58	0.86	0.01	0.06	0.01	4.47	0.33	0.73	0.63	0.82	0.01	0.06	0.01	4.6	0.37	0.72	0.65
		4	**0.93**	**0**	**0.06**	**0**	**4.2**	**0.27**	**0.58**	**0.46**	0.86	0	0.06	0.01	4.36	0.28	0.64	0.5	0.82	0.01	0.06	0.01	4.53	0.28	0.66	0.5
	5	0	0.9	0	0.06	0	4.57	0.33	0.63	0.36	0.82	0.01	0.07	0.01	4.71	0.34	0.65	0.34	0.79	0.01	0.07	0.01	4.89	0.36	0.68	0.37
		1	0.91	0	0.06	0	4.43	0.3	0.54	0.39	0.84	0.01	0.06	0	4.55	0.33	0.57	0.45	0.8	0.01	0.07	0.01	4.68	0.32	0.62	0.48
		2	0.92	0	0.06	0.01	4.51	0.3	0.72	0.54	0.84	0.01	0.07	0.01	4.65	0.37	0.73	0.54	0.81	0.01	0.07	0.01	4.78	0.36	0.79	0.52
		3	0.93	0	0.06	0	4.41	0.24	0.74	0.47	0.85	0	0.06	0.01	4.59	0.28	0.84	0.52	0.82	0.01	0.07	0.01	4.74	0.3	0.9	0.57
		4	**0.93**	**0**	**0.06**	**0.01**	**4.36**	**0.41**	**0.59**	**0.48**	0.85	0.01	0.06	0.01	4.55	0.37	0.66	0.49	0.81	0.01	0.07	0.01	4.72	0.36	0.68	0.57

Con respecto a la ciudad de Quito tenemos: el número de muestras idóneo para este caso de estudio fueron 5 para todos los días, con la diferencia en el número de retrasos teniendo como resultado que para el día 1 el mejor resultado se obtuvo con el retraso de 3 con un R^2=0,95, MAPE=0,05, RMSE=3,56 y PBIAS=0,59; para el día 2 el retraso adecuado fue de 2 con un R^2=0,94, MAPE=0,05, RMSE=3,83 y PBIAS=0,55 siguiendo con el día 3 el cual tuvo el retraso de 1 con un R^2=0,93, MAPE=0,06, RMSE=4 y PBIAS=0,54, para el día 4 y 5 el mejor retraso fue de 4 con un R^2=0,93, MAPE=0,06, RMSE=4,2 y PBIAS=0,58 y R^2=0,93, MAPE=0,06, RMSE=4,36 y PBIAS=0,59. Finalmente, se puede decir que la disminución del número de retraso influye en la disminución de las métricas de rendimiento.

6.5.1.4. Predicción del NO_2 mediante Random Forest considerando el número mínimo de muestras en los nodos hoja.

Tabla 27. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para NO_2 de Cuenca.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Cuenca	1	0	0.9	0	0.08	0.01	0.51	0.03	1.01	0.65	0.83	0.01	0.08	0.01	0.5	0.03	1.01	0.67	0.81	0.01	0.08	0.01	0.5	0.03	0.99	0.68
		1	**0.91**	**0**	**0.08**	**0.01**	**0.5**	**0.03**	**0.8**	**0.65**	0.85	0	0.08	0.01	0.5	0.03	0.79	0.63	0.82	0.01	0.08	0.01	0.5	0.03	0.82	0.61
		2	0.92	0	0.09	0.01	0.5	0.04	1.01	0.73	0.86	0.01	0.09	0.01	0.5	0.04	0.97	0.74	0.83	0.01	0.09	0.01	0.5	0.04	0.97	0.75
		3	0.93	0	0.09	0.01	0.5	0.04	0.83	0.58	0.86	0.01	0.09	0.01	0.5	0.03	0.83	0.58	0.83	0.01	0.09	0.01	0.5	0.04	0.83	0.57
		4	0.93	0	0.09	0.01	0.5	0.04	0.96	0.74	0.85	0.01	0.09	0.01	0.5	0.04	0.95	0.73	0.82	0.01	0.09	0.01	0.5	0.04	0.97	0.72
	2	0	0.89	0	0.09	0.01	0.55	0.03	1.1	0.76	0.81	0.01	0.09	0.01	0.54	0.03	1.12	0.76	0.79	0.01	0.09	0.01	0.54	0.03	1.1	0.78
		1	**0.9**	**0**	**0.09**	**0.01**	**0.53**	**0.04**	**0.76**	**0.58**	0.83	0.01	0.09	0.01	0.53	0.04	0.74	0.57	0.8	0.01	0.09	0.01	0.53	0.04	0.76	0.6
		2	0.91	0	0.09	0.01	0.53	0.03	0.81	0.83	0.83	0.01	0.09	0.01	0.53	0.03	0.87	0.81	0.81	0.01	0.09	0.01	0.53	0.03	0.85	0.81
		3	0.91	0	0.09	0.01	0.52	0.04	0.96	0.78	0.83	0.01	0.09	0.01	0.52	0.04	0.93	0.77	0.8	0.01	0.09	0.01	0.52	0.04	0.96	0.8
		4	0.92	0	0.09	0.01	0.53	0.03	1.03	0.84	0.84	0.01	0.09	0.01	0.53	0.03	1.03	0.77	0.8	0.01	0.09	0.01	0.53	0.03	1.04	0.84
	3	0	**0.87**	**0**	**0.1**	**0.01**	**0.56**	**0.04**	**0.72**	**0.5**	0.79	0.01	0.1	0.01	0.56	0.04	0.72	0.49	0.77	0.01	0.1	0.01	0.56	0.04	0.73	0.45
		1	0.89	0	0.1	0.01	0.57	0.04	0.8	0.62	0.81	0.01	0.1	0.01	0.57	0.04	0.8	0.6	0.78	0.01	0.1	0.01	0.57	0.04	0.8	0.6
		2	0.9	0	0.1	0.01	0.57	0.04	1	0.88	0.82	0.01	0.1	0.01	0.56	0.04	1	0.84	0.79	0.01	0.1	0.01	0.56	0.05	1	0.86
		3	0.91	0	0.1	0.01	0.55	0.03	1.08	0.89	0.82	0.01	0.1	0.01	0.56	0.03	1.04	0.91	0.79	0.01	0.1	0.01	0.56	0.03	1.02	0.94
		4	0.91	0.01	0.1	0.01	0.55	0.04	1.26	0.96	0.82	0.01	0.1	0.01	0.56	0.04	1.25	0.94	0.78	0.01	0.1	0.01	0.56	0.03	1.3	0.94
	4	0	0.86	0.01	0.11	0.01	0.61	0.04	0.92	0.75	0.77	0.01	0.1	0.01	0.6	0.04	0.95	0.72	0.74	0.01	0.1	0.01	0.6	0.04	0.93	0.72
		1	0.88	0	0.1	0.01	0.59	0.04	1.14	0.92	0.79	0.01	0.1	0.01	0.58	0.04	1.08	0.9	0.76	0.01	0.1	0.01	0.59	0.04	1.06	0.9
		2	0.9	0	0.1	0.01	0.58	0.03	1.14	0.9	0.81	0.01	0.1	0.01	0.58	0.03	1.16	0.93	0.77	0.01	0.1	0.01	0.59	0.03	1.18	0.98
		3	**0.9**	**0**	**0.1**	**0.01**	**0.56**	**0.04**	**0.87**	**0.57**	0.81	0.01	0.1	0.01	0.56	0.04	0.89	0.57	0.77	0.01	0.1	0.01	0.57	0.04	0.92	0.57
		4	0.91	0.01	0.1	0.01	0.55	0.05	1.04	0.82	0.82	0.01	0.1	0.01	0.56	0.05	1.05	0.85	0.77	0.01	0.1	0.01	0.56	0.05	1.06	0.86
	5	0	0.86	0	0.1	0.01	0.6	0.03	0.87	0.71	0.76	0.01	0.1	0.01	0.6	0.03	0.87	0.75	0.73	0.01	0.1	0.01	0.6	0.03	0.87	0.77
		1	0.89	0	0.1	0.01	0.57	0.02	1.11	0.62	0.79	0.01	0.1	0.01	0.57	0.03	1.12	0.62	0.76	0.01	0.1	0.01	0.58	0.03	1.14	0.66
		2	0.9	0	0.1	0.01	0.57	0.03	1.28	1.07	0.8	0.01	0.1	0.01	0.58	0.03	1.3	1.1	0.77	0.01	0.1	0.01	0.58	0.03	1.36	1.11
		3	**0.91**	**0**	**0.1**	**0.01**	**0.55**	**0.03**	**0.81**	**0.59**	0.81	0.01	0.1	0.01	0.56	0.04	0.85	0.63	0.77	0.01	0.1	0.01	0.57	0.04	0.78	0.63
		4	0.91	0	0.09	0.01	0.54	0.04	1.15	0.79	0.81	0.01	0.1	0.01	0.55	0.04	1.15	0.9	0.77	0.01	0.1	0.01	0.56	0.04	1.14	0.92

Con respecto al contaminante NO2 para el caso de Cuenca se puede ver un notable resultado óptimo con relación al número mínimo de muestras de 5. Sin embargo, muestra diferencias en el número de retrasos o variables ficticias creadas teniendo como resultado para el día 1 y 2 el retraso de 1 con una R^2=0,91, MAPE=0,08, RMSE=0,5 y PBIAS=0,8 y R^2=0,9, MAPE=0,09, RMSE=0,53 y PBIAS=0,76 en el orden dado, para el día 3 el retraso fue de 0 con un R^2=0,87, MAPE=0,1, RMSE=0,56 y PBIAS=0,72 y para los días 4 y 5 el mejor resultado obtenido fue con el retraso 3 con un R^2=0,9, MAPE=0,1, RMSE=0,56 y PBIAS=0,87 y R^2=0,91, MAPE=0,1, RMSE=30,55 y PBIAS=0,81, lo que muestra que mientras aumentó el número de días a predecir aumentó el valor de retrasos.

Tabla 28. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para NO_2 de Guayaquil.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Guayaquil	1	0	0.9	0	0.06	0	0.68	0.05	0.63	0.45	0.84	0.01	0.06	0	0.67	0.05	0.62	0.47	0.82	0.01	0.06	0	0.67	0.05	0.64	0.47
		1	0.92	0	0.06	0	0.63	0.04	0.75	0.65	0.87	0.01	0.06	0	0.63	0.04	0.71	0.64	0.84	0.01	0.06	0	0.63	0.04	0.69	0.61
		2	0.93	0	0.06	0	0.63	0.04	0.8	0.59	0.88	0.01	0.06	0	0.63	0.04	0.83	0.56	0.86	0.01	0.06	0	0.64	0.04	0.81	0.55
		3	**0.94**	**0**	**0.06**	**0**	**0.59**	**0.05**	**0.63**	**0.42**	0.89	0.01	0.06	0	0.59	0.05	0.66	0.4	0.87	0.01	0.06	0	0.6	0.05	0.69	0.42
		4	0.94	0	0.06	0	0.58	0.04	0.75	0.52	0.89	0	0.06	0	0.58	0.04	0.76	0.49	0.87	0.01	0.06	0.01	0.59	0.05	0.77	0.52
	2	0	0.88	0.01	0.07	0	0.76	0.06	0.8	0.59	0.8	0.01	0.07	0	0.76	0.06	0.8	0.61	0.77	0.01	0.07	0	0.77	0.06	0.8	0.65
		1	0.91	0	0.06	0	0.72	0.05	0.76	0.66	0.83	0.01	0.06	0	0.72	0.05	0.77	0.65	0.81	0.01	0.07	0	0.73	0.05	0.76	0.68
		2	0.92	0	0.06	0	0.68	0.05	0.79	0.65	0.86	0.01	0.06	0	0.68	0.05	0.76	0.62	0.83	0.01	0.06	0	0.68	0.04	0.81	0.64
		3	**0.93**	**0**	**0.06**	**0**	**0.69**	**0.04**	**0.6**	**0.52**	0.86	0.01	0.06	0	0.69	0.05	0.58	0.53	0.83	0.01	0.07	0	0.71	0.05	0.6	0.53
		4	0.93	0	0.06	0.01	0.66	0.06	0.93	0.61	0.86	0.01	0.06	0.01	0.67	0.06	0.93	0.62	0.83	0.01	0.06	0.01	0.69	0.06	0.95	0.65
	3	0	0.87	0	0.07	0	0.79	0.05	0.9	0.52	0.79	0.01	0.07	0	0.79	0.05	0.91	0.49	0.76	0.01	0.07	0	0.8	0.05	0.95	0.51
		1	0.9	0	0.07	0	0.77	0.04	0.85	0.67	0.81	0.01	0.07	0	0.77	0.04	0.84	0.7	0.78	0.01	0.07	0	0.78	0.04	0.86	0.7
		2	**0.91**	**0**	**0.07**	**0**	**0.75**	**0.05**	**0.85**	**0.56**	0.83	0.01	0.07	0	0.75	0.05	0.86	0.52	0.79	0.01	0.07	0	0.76	0.05	0.9	0.57
		3	0.92	0	0.07	0	0.74	0.05	0.8	0.6	0.84	0.01	0.07	0	0.75	0.05	0.79	0.63	0.8	0.01	0.07	0	0.77	0.05	0.83	0.61
		4	0.92	0	0.07	0.01	0.74	0.06	1.01	0.88	0.84	0.01	0.07	0.01	0.75	0.06	1.07	0.9	0.8	0.01	0.07	0.01	0.77	0.06	1.09	0.97
	4	0	0.86	0	0.08	0	0.84	0.05	0.7	0.59	0.76	0.01	0.08	0	0.84	0.05	0.71	0.57	0.72	0.01	0.08	0	0.85	0.05	0.74	0.57
		1	0.88	0	0.08	0	0.83	0.05	0.81	0.65	0.79	0.01	0.08	0.01	0.83	0.05	0.84	0.66	0.75	0.01	0.08	0.01	0.84	0.05	0.89	0.66
		2	0.9	0	0.08	0	0.8	0.05	0.9	0.63	0.81	0.01	0.08	0	0.8	0.05	0.95	0.59	0.77	0.01	0.08	0.01	0.82	0.05	1.03	0.57
		3	**0.91**	**0**	**0.08**	**0.01**	**0.78**	**0.05**	**0.78**	**0.62**	0.82	0.01	0.08	0.01	0.79	0.05	0.81	0.7	0.78	0.01	0.08	0.01	0.82	0.05	0.88	0.76
		4	0.92	0	0.07	0.01	0.76	0.05	0.93	0.65	0.82	0.01	0.08	0.01	0.77	0.05	0.94	0.63	0.78	0.01	0.08	0.01	0.8	0.05	0.91	0.67
	5	0	0.85	0.01	0.08	0.01	0.86	0.05	0.7	0.58	0.74	0.01	0.08	0.01	0.86	0.05	0.7	0.59	0.71	0.01	0.08	0.01	0.87	0.05	0.7	0.58
		1	0.88	0.01	0.08	0	0.83	0.04	0.81	0.8	0.78	0.01	0.08	0	0.84	0.04	0.81	0.77	0.74	0.01	0.08	0	0.85	0.04	0.85	0.78
		2	0.9	0	0.08	0.01	0.82	0.05	0.81	0.57	0.8	0.01	0.08	0.01	0.83	0.05	0.84	0.57	0.76	0.01	0.08	0.01	0.85	0.05	0.88	0.58
		3	0.91	0	0.08	0.01	0.8	0.05	0.82	0.62	0.82	0.01	0.08	0.01	0.81	0.05	0.83	0.67	0.77	0.01	0.08	0.01	0.83	0.05	0.87	0.71
		4	**0.92**	**0**	**0.07**	**0.01**	**0.77**	**0.06**	**0.89**	**0.64**	0.82	0.01	0.08	0.01	0.79	0.06	0.93	0.66	0.77	0.01	0.08	0.01	0.83	0.06	1	0.73

De la misma manera para Guayaquil muestra que para predecir el día 1 se optó como mejor estructura con retraso de 3 con un R^2=0,94, MAPE=0,06, RMSE=0,59 y PBIAS=0,63, seguido de un retraso de 3 para el día 2 con un R^2=0,93, MAPE=0,07, RMSE=0,75 y PBIAS=0,85, en el día 3 el retraso disminuye mostrando un valor de 3 con un R^2=0.91, MAPE=0,07, RMSE=0,75 y PBIAS=0,85, el día 4 se puede observar como mejor resultado el retraso de 3 con un R^2=0,91, MAPE=0,08, RMSE=0,78 y PBIAS=0,78, finalmente para predecir a 5 días hacia adelante la mejor respuesta fue un retraso de 4 con un R^2=0,92, MAPE=0,07, RMSE=0,77 y PBIAS=0,89, teniendo en cuenta que el número mínimo de muestras fue de 5 para todos los días y la bondad de ajuste va disminuyendo mientras aumenta el valor del mismo.

Tabla 29. Resultados de análisis de Random Forest tomando en cuenta el número mínimo de muestras en los nodos hoja para NO_2 de Quito.

			Mínimo de muestras 5								Mínimo de muestras 15								Mínimo de muestras 25							
	Día adelante	Retrasos	R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS		R2		MAPE		RMSE		PBIAS	
Quito	1	0	0.89	0.01	0.07	0	0.81	0.05	0.85	0.55	0.82	0.01	0.07	0	0.79	0.05	0.8	0.55	0.8	0.01	0.07	0	0.79	0.05	0.76	0.53
		1	**0.91**	**0**	**0.07**	**0.01**	**0.77**	**0.05**	**0.77**	**0.59**	0.84	0.01	0.07	0.01	0.76	0.05	0.73	0.59	0.81	0.01	0.07	0.01	0.78	0.05	0.7	0.61
		2	0.92	0	0.07	0.01	0.8	0.05	0.75	0.53	0.84	0.01	0.07	0.01	0.79	0.05	0.7	0.54	0.81	0.01	0.07	0.01	0.81	0.06	0.73	0.59
		3	0.92	0	0.07	0.01	0.78	0.07	0.97	0.64	0.85	0.01	0.07	0.01	0.77	0.07	0.93	0.69	0.81	0.01	0.07	0.01	0.8	0.06	0.95	0.74
		4	0.92	0	0.08	0.01	0.81	0.07	1	0.72	0.85	0.01	0.08	0.01	0.81	0.07	0.98	0.73	0.8	0.01	0.08	0.01	0.85	0.08	0.99	0.73
	2	0	0.86	0	0.08	0.01	0.93	0.07	0.84	0.69	0.77	0.01	0.08	0.01	0.92	0.07	0.81	0.63	0.73	0.01	0.08	0.01	0.93	0.07	0.81	0.59
		1	0.87	0.01	0.08	0.01	0.9	0.06	0.68	0.63	0.78	0.01	0.08	0.01	0.9	0.06	0.65	0.57	0.74	0.01	0.08	0.01	0.9	0.06	0.64	0.57
		2	0.89	0.01	0.08	0.01	0.91	0.07	1.13	0.74	0.79	0.01	0.08	0.01	0.91	0.07	1.08	0.74	0.75	0.01	0.09	0.01	0.92	0.07	1.01	0.72
		3	0.9	0	0.08	0.01	0.88	0.06	1.02	0.71	0.8	0.01	0.08	0.01	0.89	0.06	1.03	0.74	0.76	0.01	0.09	0.01	0.91	0.07	1.06	0.74
		4	**0.91**	**0**	**0.08**	**0.01**	**0.86**	**0.07**	**0.86**	**0.61**	0.81	0.01	0.08	0.01	0.87	0.07	0.83	0.67	0.75	0.01	0.08	0.01	0.9	0.07	0.88	0.65
	3	0	0.83	0.01	0.09	0.01	1.01	0.06	1.06	0.75	0.73	0.01	0.09	0.01	1	0.06	1.02	0.69	0.69	0.01	0.09	0.01	1	0.06	1.02	0.7
		1	0.86	0.01	0.09	0.01	0.96	0.05	0.93	0.65	0.75	0.01	0.09	0.01	0.96	0.05	0.93	0.64	0.71	0.01	0.09	0.01	0.97	0.05	0.91	0.66
		2	0.88	0.01	0.09	0.01	0.95	0.07	1.18	0.68	0.77	0.01	0.09	0.01	0.95	0.06	1.18	0.7	0.72	0.01	0.09	0.01	0.96	0.06	1.14	0.75
		3	0.89	0	0.09	0.01	0.93	0.07	1.09	0.77	0.78	0.01	0.09	0.01	0.95	0.06	1.1	0.78	0.72	0.01	0.09	0.01	0.97	0.06	1.16	0.9
		4	**0.9**	**0.01**	**0.09**	**0.01**	**0.93**	**0.06**	**1.05**	**0.67**	0.79	0.01	0.09	0.01	0.94	0.06	1.13	0.77	0.72	0.02	0.09	0.01	0.97	0.06	1.23	0.79
	4	0	0.83	0.01	0.09	0.01	1.01	0.05	0.91	0.7	0.72	0.01	0.09	0.01	1	0.05	0.94	0.72	0.68	0.01	0.09	0.01	1	0.05	0.95	0.75
		1	0.86	0.01	0.09	0.01	0.98	0.07	1.13	0.78	0.75	0.01	0.09	0.01	0.97	0.07	1.1	0.75	0.7	0.01	0.09	0.01	0.99	0.07	1.1	0.76
		2	0.88	0.01	0.09	0.01	0.97	0.08	1.21	0.88	0.77	0.01	0.09	0.01	0.97	0.07	1.23	0.9	0.72	0.01	0.09	0.01	0.98	0.07	1.22	0.93
		3	0.89	0	0.09	0.01	0.95	0.05	0.97	0.76	0.78	0.01	0.09	0.01	0.97	0.05	1.01	0.77	0.72	0.01	0.09	0.01	0.99	0.06	1.08	0.77
		4	**0.9**	**0.01**	**0.09**	**0.01**	**0.93**	**0.07**	**1.01**	**0.72**	0.79	0.01	0.09	0.01	0.95	0.07	1.1	0.79	0.73	0.02	0.09	0.01	0.99	0.07	1.22	0.88
	5	0	0.83	0.01	0.09	0.01	0.98	0.06	0.67	0.58	0.72	0.01	0.09	0.01	0.98	0.07	0.7	0.64	0.68	0.01	0.09	0.01	0.99	0.07	0.75	0.66
		1	0.87	0.01	0.09	0.01	0.95	0.06	0.9	0.64	0.75	0.01	0.09	0.01	0.95	0.06	0.89	0.65	0.71	0.01	0.09	0.01	0.97	0.06	0.97	0.69
		2	0.88	0.01	0.09	0.01	0.96	0.06	1	0.75	0.77	0.01	0.09	0.01	0.97	0.06	1.01	0.77	0.72	0.01	0.09	0.01	0.99	0.05	1.01	0.83
		3	0.89	0	0.09	0.01	0.93	0.06	0.95	0.67	0.78	0.01	0.09	0.01	0.95	0.06	0.99	0.68	0.73	0.01	0.09	0.01	0.98	0.07	1.05	0.7
		4	**0.9**	**0.01**	**0.09**	**0.01**	**0.91**	**0.06**	**0.94**	**0.73**	0.79	0.01	0.09	0.01	0.94	0.06	1.03	0.76	0.73	0.01	0.09	0.01	0.97	0.06	1.12	0.86

Quito muestras resultados similares a Cuenca y Guayaquil mostrando cierta tendencia en el modelo donde el número de muestras fue de 5 lo que muestra la extensión ideal para cada árbol de decisión, para el caso del número de retraso para cada día a predecir podemos decir que en su mayoría el valor ideal fue de 4 exceptuando para el primer día donde el mejor retraso es de 1. Con respecto a las métricas de rendimiento podemos decir que la mejor predicción lograda es para el día 1 con un R^2=0,91, MAPE=0,07, RMSE=0,77 y PBIAS=0,77,teniendo en cuenta el PBIAS ya que el día 2 tienen resultado similares con un R^2=0,91, MAPE=0,08, RMSE=0,86 y PBIAS=0,86 seguido por los días 5, 4 y 3 con valores de R^2=0,9, MAPE=0,09, RMSE=0,91 y PBIAS=0,94; R^2=0,9, MAPE=0,09, RMSE=0,93 y PBIAS=1,01; y R^2=0,9, MAPE=0,09, RMSE=0,93 y PBIAS=1,05 respectivamente.

6.6. Tablas resumen de los las Predicciones mediante RNN Y Random Forest.

Tabla 30. Tabla Resumen de Predicciones de NO_2 y SO_2 mediante Redes Neuronales Recurrentes.

Tabla de predicción mediante Redes Neuronales Recurrentes

Redes Neuronales Recurrentes para SO2

	Cuenca						Guayaquil						Quito					
Días	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*	**Días**	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*	**Días**	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*
1	0	64	0.96	7.7	4.06	0.70	**1**	0	64	0.87	4.7	8.49	0.55	**1**	1	64	0.86	5.9	4.33	0.55
2	0	64	0.95	8.3	0.08	0.08	**2**	1	64	0.85	5.9	10.70	0.69	**2**	1	64	0.85	6.2	4.42	0.66
3	0	32	0.94	9.6	5.23	0.80	**3**	1	64	0.83	6.3	11.34	0.83	**3**	1	64	0.84	6.4	4.50	0.50
4	0	64	0.92	10.5	5.84	0.74	**4**	1	64	0.80	6.5	11.86	0.83	**4**	0	64	0.81	6.5	4.54	0.66
5	0	64	0.91	11.2	6.19	0.68	**5**	1	64	0.79	7.0	12.56	0.77	**5**	1	64	0.82	6.9	4.87	0.71

Redes Neuronales Recurrentes para NO2

	Cuenca						Guayaquil						Quito					
Días	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*	**Días**	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*	**Días**	*Retrasos*	*# Neuronas*	*R2*	*MAPE*	*RMSE*	*PBIAS*
1	0	64	0.811	9.1	0.54	0.978	**1**	1	64	0.85	7.3	0.79	0.81	**1**	0	64	0.8	7	0.814	0.723
2	1	64	0.816	10.7	0.63	0.971	**2**	1	64	0.83	7.8	0.856	0.946	**2**	1	64	0.74	9.4	1.023	0.924
3	0	32	0.77	10.0	0.59	0.814	**3**	1	64	0.79	8.7	0.943	1.091	**3**	2	64	0.77	11	1.164	1.057
4	1	64	0.767	11.8	0.69	1.466	**4**	1	64	0.77	9.0	0.957	0.841	**4**	2	64	0.78	11	1.173	1.212
5	1	64	0.768	11.5	0.67	1.495	**5**	1	64	0.75	9.3	0.983	1.137	**5**	2	64	0.78	11	1.181	1.217

Un análisis general de predicción mediante redes neuronales recurrentes para NO_2 y SO_2 se presenta en la tabla 30 donde podemos apreciar los mejores resultados de cada arquitectura y cada área de estudio. Las métricas de bondad de ajuste y error son mejores para las emisiones de dióxido de Azufre SO_2 en comparación con las concentraciones de dióxido de nitrógeno NO_2.

Tabla 31. Tabla Resumen de predicciones de NO_2 y SO_2 mediante Random Forest.

Tabla de predicción mediante Random Forest

Random Forest para SO2 con profundidad máxima

	Cuenca							Guayaquil							Quito					
Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS
1	2	15	0.98	6.2	3.3	0.43	1	3	25	0.98	4	7.19	0.43	1	3	25	0.98	5	3.55	0.57
2	3	15	0.98	6.5	3.43	0.72	2	3	15	0.98	5	8.28	0.54	2	3	25	0.98	5	3.78	0.65
3	2	25	0.98	7	3.7	0.7	3	2	15	0.97	5	9.37	0.59	3	3	25	0.98	5	3.99	0.77
4	1	25	0.98	7.3	3.94	0.92	4	3	15	0.97	6	10.01	0.55	4	4	25	0.97	6	4.12	0.59
5	4	15	0.98	7.8	4	0.63	5	4	15	0.96	6	10.51	0.69	5	3	25	0.97	6	4.37	0.71

Random Forest para NO2 con profundidad máxima

	Cuenca							Guayaquil							Quito					
Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Prof.	R2	MAPE	RMSE	PBIAS
1	1	15	0.97	8.4	0.5	0.75	1	3	15	0.98	5.6	0.593	0.62	1	1	15	0.97	7	0.78	0.798
2	3	15	0.96	9.1	0.52	1	2	3	15	0.97	6.4	0.688	0.59	2	4	15	0.96	8	0.86	0.843
3	3	15	0.96	9.6	0.56	1.09	3	3	15	0.97	7	0.74	0.79	3	4	15	0.96	9	0.92	0.983
4	3	25	0.96	9.8	0.56	0.87	4	3	15	0.96	7.5	0.782	0.82	4	4	15	0.96	9	0.93	0.957
5	3	15	0.96	9.6	0.55	0.74	5	4	15	0.96	7.3	0.759	0.87	5	4	15	0.96	9	0.91	0.915

Random Forest para SO2 con número mínimo de muestras

	Cuenca							Guayaquil							Quito					
Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS
1	2	5	0.96	6	3.27	0.44	1	3	5	0.95	4	7.14	0.44	1	3	5	0.95	5	3.56	0.59
2	2	5	0.95	7	3.52	0.65	2	3	5	0.95	4	7.14	0.44	2	2	5	0.94	5	3.83	0.55
3	2	5	0.95	7	3.69	0.61	3	2	5	0.93	5	9.37	0.61	3	1	5	0.93	6	4	0.54
4	3	5	0.94	7	3.85	0.75	4	3	5	0.92	6	10.08	0.57	4	4	5	0.93	6	4.2	0.58
5	4	5	0.94	8	4.04	0.65	5	4	5	0.91	6	10.62	0.72	5	4	5	0.93	6	4.36	0.59

Random Forest para NO2 con número mínimo de muestras

	Cuenca							Guayaquil							Quito					
Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS	Días	Retrasos	Muestra	R2	MAPE	RMSE	PBIAS
1	1	5	0.91	8.4	0.5	0.81	1	3	5	0.94	5.6	0.594	0.63	1	1	5	0.91	7	0.77	0.772
2	1	5	0.9	9.3	0.53	0.76	2	3	5	0.93	6.4	0.686	0.6	2	4	5	0.91	8	0.86	0.857
3	0	5	0.87	9.6	0.56	0.72	3	2	5	0.91	7	0.751	0.85	3	4	5	0.9	9	0.93	1.051
4	3	5	0.9	9.8	0.56	0.87	4	3	5	0.91	7.5	0.783	0.78	4	4	5	0.9	9	0.93	1.012
5	3	5	0.91	9.6	0.55	0.81	5	4	5	0.92	7.4	0.766	0.89	5	4	5	0.9	9	0.91	0.942

Un resumen de los resultados de predicción de SO_2 y NO_2 mediante Random Forest se muestra en la tabla 31 tomando en cuenta dos características principales observando mejores resultados con el análisis de profundidad máxima con la diferencia que la mejor estructura se realiza con mayores valores de profundidad. En el caso del número mínimo de muestras se obtienen mejores resultados con menos cantidad de muestras. Además, al igual que en el primer método, la predicción de Dióxido de Azufre SO_2 fueron mejores a diferencia de Dióxido de Nitrógeno NO_2.

Tabla 32. Tabla de promedios de resultados de Predicción de NO_2 y SO_2.

Promedios de Resultados de Predicción de Contaminantes atmosféricos							
		SO_2			*NO_2*		
	Métricas	Cuenca	Guayaquil	Quito	Cuenca	Guayaquil	Quito
Redes Neuronales Recurrentes	R2	0,94	0,83	0,84	0,79	0,80	0,78
	MAPE	9,5	6,1	6,4	10,6	8,4	9,9
	RMSE	4,28	10,99	4,53	0,62	0,91	1,07
	PBIAS	0,60	0,73	0,62	1,14	0,97	1,03
RF con Profundidad máxima	R2	0,98	0,97	0,97	0,96	0,97	0,96
	MAPE	7,0	5,2	5,4	9,3	6,8	8,4
	RMSE	3,67	9,07	3,96	0,53	0,71	0,88
	PBIAS	0,68	0,56	0,65	0,88	0,74	0,89
RF con Número mínimo de muestras	R2	0,94	0,93	0,936	0,89	0,92	0,90
	MAPE	7	5	6	9,3	6,8	8,4
	RMSE	3,67	8,87	3,99	0,54	0,71	0,88
	PBIAS	0,62	0,556	0,57	0,794	0,74	0,92

En la tabla número 33 se exponen los valores obtenidos para la predicción de NO_2 y SO_2 para las ciudades de Cuenca, Guayaquil y Quito. Como premisa se puede observar que, aunque los dos métodos de predicción generan valores cercanos al original el modelo que representa una predicción más cercana al valor original es el método de Random Forest.

Tabla 33. Pronóstico para 5 días hacia adelante de NO_2 y SO_2 para el año 2021.

Pronóstico para 5 días hacia adelante mediante RF y RNN										
		Cuenca			***Guayaquil***			***Quito***		
	Fecha	*Real*	*RNN*	*RF*	*Real*	*RNN*	*RF*	*Real*	*RNN*	*RF*
NO_2	1/1/21	3,6986	4,0420	3,5347	6,3061	9,9139	6,6530	6,6791	8,6767	6,5225
	2/1/21	3,3398	4,8925	3,4615	6,8099	10,8461	6,1716	6,8918	11,2338	7,3676
	3/1/21	3,0406	4,6761	3,1951	6,3680	10,6994	6,5337	7,6256	11,3680	7,6807
	4/1/21	3,6021	4,9403	3,7077	7,0871	9,4672	7,1964	7,7915	13,2240	7,9297
	5/1/21	3,4872	5,1107	3,6693	7,2668	9,9911	7,3410	7,9976	12,9577	8,0561
		Cuenca			***Guayaquil***			***Quito***		
	Fecha	*Real*	*RNN*	*RF*	*Real*	*RNN*	*RF*	*Real*	*RNN*	*RF*
SO_2	1/1/21	28,6847	25,0780	26.4995	96,1041	96,4674	96,3325	35,2023	49,4651	37,0975
	2/1/21	27,1228	28,6612	26.1122	108,3017	109,2421	109,0663	35,5011	39,7693	34,0704
	3/1/21	24,7937	32,3622	25.7618	111,1305	124,2552	111,4256	35,6069	44,9950	35,1398
	4/1/21	26,2448	28,7057	26.3492	122,6756	118,0854	122,0668	36,1923	41,4295	37,3515
	5/1/21	25,9051	36,1852	25.7212	105,9549	99,9781	106,0757	37,2560	45,1529	38,3305

7. DISCUSIÓN

7.1. Análisis de la predicción de concentraciones de NO_2 y SO_2 mediante Redes Neuronales Recurrentes

Actualmente la contaminación del aire ambiente representa un factor importante para el desarrollo de la sociedad. Es por esto que realizar métodos de predicción de contaminantes utilizando técnicas de Machine Learning es uno de los procesos que necesariamente deben ser estudiados con mayor énfasis, esto debido a las ventajas obtenidas con el avance tecnológico, en donde en concordancia con el estudio realizado por (Rybarczyk y Zalakeviciute, 2018) se afronta las limitaciones que se dan por realizar modelos deterministas tradicionales, los cuales al no poder relacionar los datos no lineales que se dan entre la concentración de contaminantes del aire y sus fuentes de emisión. Propone la utilización de modelos estadísticos basados en Machine Learning en donde como principal hallazgo se determina que el aprendizaje automático es aplicado principalmente en países con un elevado nivel tecnológico siendo Europa y América del norte los continentes que lo aplican principalmente y que para realizar pronósticos se pueden utilizar modelos de redes neuronales y máquinas de vectores de soporte.

Para realizar el método de predicción mediante redes neuronales recurrentes se estimó que para la arquitectura de la misma lo óptimo sería utilizar una capa oculta. En donde con base en esto se decidió probar con 16, 32 y 64 neuronas. Siendo que para el conjunto de datos estudiado, el número óptimo de neurona para todos los eventos fue de 64 neuronas.

En el documento realizado por (Rani Hemamalini et al., 2022) se desea realizar un monitoreo mediante la utilización de drones los cuales se elevan a 100 metros del nivel del suelo y un pronóstico de la calidad del aire aplicando redes neuronales recurrentes. Lo que se realiza es adaptar a drones sensores los cuales sean capaces de determinar la concentración de varios contaminantes entre ellos el NO_2 y el SO_2 para analizar dos zonas, la primera se encuentra en los alrededores de un botadero de residuos sólidos y la segunda son las áreas residenciales industriales del distrito de Manali en Chennai, India. Por practicidad tomó en cuenta sólo la zona residencial y las mediciones se realizaron entre enero del 2021 y febrero de 2022. Posteriormente se utiliza un modelo de redes neuronales recurrente bidireccional para predecir la concentración de contaminantes Para la zona residencial el modelo propuesto predice con precisión las concentraciones de los

contaminantes , esto debido a que obtuvo valores bajos en las métricas de evaluación de errores, para este documento se analizó el RMSE, MAE y el MAPE, sin embargo para comparar los resultados solo se tomará en cuenta 2 métricas. El RMSE promedio en la zona residencial es de 1,35 y el MAPE promedio es de 1,53. Los valores promedios de RMSE obtenidos con nuestro modelo para el NO_2 son para Cuenca de 0,62, para Guayaquil es de 0,91 y para Quito es de 1,07. Para el SO_2 el valor promedio en Cuenca es de 4,28, para Guayaquil es de 10,99, para Quito es de 4,53. El MAPE promedio obtenido en nuestro modelo para el SO_2 en Cuenca es de 9,5, para Guayaquil es de 6,1 y para Quito es de 6,4. Para el NO_2 el MAPE promedio en cuenca nos resultó en un valor de 10,6, para Guayaquil es de 8,4 y para Quito el valor es de 9,9. Debido a que el modelo se basó en sensores satelitales podemos concordar que nuestro análisis representa cierta similitud con el modelo propuesto en el documento antes mencionado, sin embargo el error obtenido en general es más elevado que el del modelo analizado.

En el documento de (Brunelli et al., 2007) se realiza una predicción a dos días de varios contaminantes entre ellos el NO_2 y el SO_2 en la ciudad de Palermo, Italia. En donde como base se utiliza un método de redes neuronales recurrentes utilizando una serie temporal que va entre el 1 de enero del 2003 y el 31 de diciembre del 2004 para ocho estaciones de la ciudad mencionada, para validar el modelo lo que se hizo fue comparar los valores predichos con valores reales registrados. En el documento utilizan varios modelos estadísticos para determinar el error pero por practicidad de este documento el único a tomar en cuenta es el RMSE, el cual es determinado para las ocho estaciones siendo que para el NO2 los valores mínimos son de 4.5 y el máximo de 32.09, siendo el valor 4.5 el que posee un RMSE más adecuado según los resultados del modelo. Para el SO2 tenemos que el valor máximo de RMSE es de 3.13 y el mínimo es de 0.58, siendo este el más adecuado. En comparación los resultados de RMSE de nuestro modelo al segundo día para el SO2 fueron en Cuenca de 0.08 como valor mínimo y de 6.71 como valor máximo. Para Guayaquil los valores fueron de 10.1 como mínimo y 22.9 como máximo. Para Quito los valores fueron de 4.3 como mínimo y de 8.03 como máximo. Y los valores de NO2 fueron en Cuenca de 0.58 como mínimo y 0.89 como máximo. Para Guayaquil fueron de 0.84 como mínimo y 1.32 como máximo. Para Quito los valores fueron de 0.95 como mínimo y 1.49 como máximo. Al comparar podemos observar que los valores tienen un patrón similar para el error RMSE y tomando en cuenta el documento analizado el cual concluye que este modelo es una herramienta potencial para la predicción de los

contaminantes, podemos asumir que de la misma manera nuestro método podría ser de gran ayuda para predecir contaminantes a dos días en las grandes ciudades.

7.2. Análisis de la predicción de concentraciones de NO_2 y SO_2 mediante Random Forest

En el desarrollo de los modelos de predicción con el método de Random Forest utilizando como características principales la profundidad máxima y el número mínimo de muestras se reflejan en la tabla 30, tanto las salidas de dióxido de azufre SO2 como para el dióxido de nitrógeno NO2 con entradas meteorológicas presenta modelos altamente ajustados a la respuesta real. Además la bondad de ajuste R^2 es considerablemente alta y los errores MAPE, RMSE Y BIAS bajos en ambos modelos y para Cuenca, Guayaquil y Quito. Además en nuestro trabajo utilizamos el porcentaje de sesgo PBIAS como métrica de error para determinar el sesgo de nuestra predicción

Cabe destacar que para Random Forest se tomó en cuenta dos características, la primera, la profundidad máxima de cada árbol de decisión y la segunda, la cantidad mínima de muestras para los nodos hoja, la cual mediante el análisis de las tablas de resultados se puede decir que con la característica de profundidad máxima los resultados obtenidos son mejor a diferencia de los modelos pronosticados considerando el número mínimo de muestras para todas las áreas de estudio.

Para los modelos con la salida de dióxido de azufre y con entradas meteorológicas para la ciudad de Cuenca presenta un valor de MAPE y un PBIAS bastante bueno, el RMSE para predecir hasta cinco días hacia adelante es el más alto comparado con los demás días. Sin embargo los valores de R^2 son altos lo que hace que el modelo se ajuste adecuadamente a la salida real esperada. De la misma manera ocurre con la ciudad de Guayaquil donde en comparación con Cuenca el R^2 es menor pero el MAPE y el PBIAS arroja valores cercanos a cero, el RMSE también muestra elevarse al aumentar los días a predecir. Quito muestra resultados similares a las dos ciudades anteriores, la bondad de ajuste con valores elevados y las métricas error bajas lo que indica que el modelo es eficiente para predecir hace 5 días. Las arquitecturas utilizadas para cada predicción fueron las misma, pero con resultados diferentes para cada una, en este caso la profundidad

máxima para Cuenca y Guayaquil oscilaban entre 15 la mayor parte y 25 a diferencia de Quito donde para predecir cada día la profundidad de 25 fue la más adecuada esto puede deberse a que a pesar de que Cuenca y Quito se encuentran en la región interandina cada una cuenta con diferentes variaciones meteorológicas.

De acuerdo a los resultados expresados en la tabla 30 los valores de RMSE para Quito Guayaquil y Cuenca son 4, 10,51 y 4,37 y los valores del MAPE son 7.8, 6 y 6 respectivamente para el dióxido de azufre, en el caso del dióxido de nitrógeno tenemos valores de RMSE son 0,55, 0,593 y 0,78 y los valores de MAPE son 9,6, 7,3 y 9 en el orden dado, un estudio realizado por (Siwek y Osowski, 2016) quienes realizaron la predicción de la calidad del aire en la ciudad de Varsovia en Polonia mediante diferentes técnicas de predicción entre ella Random Forest y Redes Neuronales determinaron la predicción de PM_{10}, O_3, SO_2 y NO_2 para el siguiente día obteniendo 18,35 y 18,27 para valores de MAPE para concentraciones de SO_2 y NO_2 en el orden dado y valores de 2,79 y 3,377 para valores de RMSE para dióxido de azufre y dióxido de nitrógeno, donde los mejores resultados fueron obtenidos por este método ya que realiza la regresión e integración al mismo tiempo lo que demuestra la superioridad de tal enfoque, valores similares se encontraron en un estudio de predicción con imágenes satelitales en Europa donde concluyeron que el rendimiento relativo de los algoritmos puede diferir con el entorno del estudio, por lo tanto, se pueden hacer recomendaciones genéricas para un algoritmo o utilizar más de un algoritmo (Chen et al., 2019), tal es el caso que (Sethi y Mittal, 2019) en un estudio en Macedonia, utilizaron tres técnicas juntas con Random Forest donde se obtuvo mayor precisión del 100 % y menor tasa de error en comparación con las técnicas utilizadas individualmente. Con respecto a los niveles de concentración de SO_2 se registraron en las grandes ciudades con mayor cantidad de industrias y población, esto se ve reflejado en un estudio realizado por (Li et al., 2019) en China donde el mayor índice de contaminación por SO_2 se encuentran en las provincias más grandes y el umbral de concentración de NO_2 está estrechamente relacionada con la urbanización (Zhan et al., 2018). Sin embargo, los resultados obtenidos por (Silibello et al., 2021) en un estudio en Italia menciona que en cuanto al NO_2 los niveles en las estaciones de tráfico urbano no fueron capturados por las simulaciones debido a la resolución horizontal adoptada y las incertidumbres relacionadas con las emisiones. Un estudio en Japón demostró que la predicción de concentraciones de dióxido de nitrógeno es efectiva con un R^2=0,84 lo que ilustra la ventaja de usar este método para modelar algoritmos de predicción eficaces (Araki et al., 2018).

En el presente estudio la predicción mediante el método de Random Forest a partir de imágenes satelitales para concentraciones de NO_2 y SO_2 se obtuvieron alta eficacia y errores bajos. Además, se ha demostrado que este método es eficiente para otro tipo de contaminantes como es el caso de la predicción de concentraciones de $PM_{2,5}$ en China en el cual utilizó RF donde con un R^2=0,76 concluyeron que es un modelo computacionalmente eficiente y de alta resolución para proporcionar concentraciones históricas confiables para estudios posteriores (Huang et al., 2018). Así mismo, en Italia (Stafoggia et al., 2019) estimar las concentraciones de PM10 y PM2.5 donde se obtuvieron resultados óptimos de predicción, por su parte, en la India se implementó algoritmos de Machine Learning para predecir contaminantes como Metano (CH4), Monóxido de Carbono (CO), Ozono (O_3), PM_{10}, $PM_{2,5}$, Dióxido de Nitrógeno (NO_2) y Dióxido de Azufre (SO_2) mostrando una efectividad del 99.4 % (Sanjeev, 2021).

7.3. Análisis de la predicción de concentraciones de NO_2 y SO_2 mediante Random Forest y Redes Neuronales

Las predicciones de concentraciones diarias de dióxido de nitrógeno y dióxido de azufre mediante los dos métodos descritos anteriormente, fueron satisfactorias. Sin embargo, uno de ellos muestra mejores resultados respecto al otro, los valores promedio de bondad de ajuste (R^2) obtenidos mediante RF para predicciones de SO_2 y NO_2 fueron: 0,94 para la ciudad de Cuenca, 0,97 para Guayaquil y 0,97 para la ciudad de Quito respectivamente y para las predicciones de NO_2 para Cuenca, Guayaquil y Quito fueron: 0,96, 0,97, 0,96 respectivamente, estos se puede apreciar en la tabla 32 donde se resumen los resultados de ambos métodos en comparación con las predicciones mediante Redes Neuronales Recurrentes donde las predicciones para SO_2 fueron 0,94, 0,83 y 0,84 para las tres ciudades mencionadas en el mismo orden y las predicciones para NO_2 fueron 0,79, 0,8, 0,78 respectivamente donde se puede ver una clara diferencia entre los resultados de ambos métodos analizados. Además, las métricas de error MAPE, RMSE y PBIAS son ligeramente más altas que RF. Nuestros resultados coinciden con las expectativas ya que la incertidumbre aumenta con el periodo más largo y conduce a mayor dificultad en el pronóstico. Esto se puede apoyar en un estudio realizado por Kang et al., (2018) donde realizaron la predicción de varios contaminantes mediante redes Neuronales con varios

modelos y RF obteniendo resultados favorables para el segundo con una precisión del 81 % en comparación con las redes Neuronales que obtuvieron una precisión del 55 %. Esto debido a que este método usa múltiples arboles de decisión y reduce el sobreajuste de cada árbol mediante su combinación después del ajuste de los hiperparámetros (Ameer, et al., 2019). Sin embargo, los resultados obtenidos mediante el método de redes neuronales recurrentes son bastante eficientes y con un buen rendimiento, los mismo podemos decir en un estudio por Guo et al, (2020) donde se pronosticó con éxito las concentraciones de $PM_{2.5}$ y recomienda tomar en cuenta las actividades humanas ya que son una de las principales razones del deterioro ambiental. En definitiva, la aplicación del lenguaje de aprendizaje no supervisado ha mostrado resultados satisfactorios con un rendimiento alto y errores bajos, se puede considerar que ambos métodos son aplicables para el objetivo propuesto con la oportunidad de aplicar nuevas técnicas y algoritmos de aprendizaje abriendo puertas a futuras investigaciones dentro del Ecuador.

8. CONCLUSIONES

La aplicación de métodos de Inteligencia artificial proporciona resultados prometedores para el pronóstico de la calidad del aire. Este trabajo de investigación obtuvo datos de concentración de gases contaminantes a partir de imágenes Satelitales Sentinel-5P y datos meteorológicos a partir de imágenes satelitales de la NASA y ERA 5 durante el periodo de dos años. El estudio tuvo lugar en tres ciudades del Ecuador, en Cuenca, Guayaquil y Quito por ser las ciudades más grandes e importantes y además de que poseen elevados índices de mala calidad del aire. Como aspecto primordial una vez obtenidas las imágenes satelitales de las variables dependientes e independientes para la escala temporal utilizada, estas fueron tratadas, limpiadas, procesadas y rellenadas con el fin de obtener los datos de buena calidad y evitar vacíos los cuales afecten en el proceso de predicción.

Se realizó dos métodos para predecir la concentración de NO_2 y SO_2 en donde para cada modelo se obtuvieron buenos resultados tomando en cuenta como análisis estadístico de los modelos la bondad de ajuste R^2 y métricas de error MAPE, RMSE y PBIAS. El método de Bosque Aleatorios (FR) ofrece mejor rendimiento de las predicciones objetivo en función de tres conjuntos de datos diferentes, haciendo énfasis en la característica de profundidad de cada árbol de decisión la cual mostro una mejor relación entre los datos arrojando resultados eficientes. Todos los resultados muestran que el uso de Redes Neuronales Recurrentes produce los resultados menos satisfactorios entre todos los métodos explorados. Los resultados también confirman que los métodos de aprendizaje automático empleados en este trabajo usando datos de imágenes satelitales pueden arrojar buenos resultados y puede usarse para estudio e investigaciones posteriores. Sin embargo, la presencia de variables climáticas extremas y nubosidad puede interferir en la detección de concentraciones de contaminantes a través de estas herramientas mencionadas, ya que produce huecos para trabajar con datos diarios.

El estudio demostró que los métodos aplicados lograron ser eficientes a la hora de predecir valores de concentración de contaminantes, esto debido a que se compararon los valores predichos por el método de Random Forest y redes neuronales recurrentes con valores de imágenes satelitales de los cinco días posteriores a la escala temporal utilizada. Se compararon los valores y se observó que los valores predichos para el día uno representan una similitud más adecuada con los valores reales, mientras que para el día cinco los

valores varían más comparándolos con los valores reales. Además, este estudio también indica que el rendimiento de la predicción varía en las diferentes ciudades y regiones del Ecuador. La comparación de los resultados de los conjuntos de datos provenientes de las tres ciudades diferentes muestra mejores resultados para la predicción de concentraciones de NO_2 y SO_2 de Cuenca, donde la disminución del rendimiento a través del paso de los días a predecir es menos pronunciada que en Guayaquil y Quito.

9. RECOMENDACIONES

Por último, podemos recomendar aplicar varios algoritmos o arquitecturas en las redes neuronales para mejorar la predicción de estos contaminantes ya que la eficiencia del pronóstico también depende de los factores meteorológicos y características geográficas del área de estudio. Además, para estudios posteriores se puede combinar datos obtenidos por imágenes satelitales y datos obtenidos por estaciones de monitoreo en caso de contar con la misma, así se puede obtener una base de datos más sólida y sin huecos para trabajar constantemente, de esta manera se puede obtener datos más precisos y consecuentemente la predicción sea más eficiente.

10. BIBLIOGRAFÍA

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *IJCSI International Journal of Computer Science, 9*(3)

Ameer, S., Ali Shah, M., Khan, A., Song, H., Maple, C., UI Islam, S., & Nabeel Asghar, M. (2019). Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities. *IEEE, 7*, pp. 128325 - 128338. DOI: 10.1109/ACCESS.2019.2925082

Aguirre Basurko, E., Anta Sanz, A., R. Barrón, L. J., & Etxeberria, M. A. (2006). Relevancia de las variables meteorológicas en el diseño de un modelo de predicción de los niveles de ozono, en tiempo real, basadoen el uso de redes neuronales. *Acta de las Jornadas Científicas de la Asociación Meteorológica Española.* Obtenido de https://pub.ame-web.org/index.php/JRD/article/view/2277

Agrafiotis, D. (2014). Video Error Concealment. In M. Wu, R. Chellappa, D. R. Bull, & S. Theodoridis (Eds.), *Academic Press Library in Signal Processing: Image and Video Compression and Multimedia* (Vol. 5, pp. 295-321). Elsevier Science. https://doi.org/10.1016/B978-0-12-420149-1.00009-0

Anenberg, S. C., Bindl, M., Brauer, M., Castillo, J. J., Cavalieri, S., Duncan, B. n., Fiore, A. M., Fuller, R., Goldberg, D. L., Henze, D. K., Hess, J., Holloway, T., James, P., Jin, X., Kheirbek, I., Kinney, P. L., Liu, Y., Mohegh, A., Patz, J., … West, J. J. (2020). Using Satellites to Track Indicators of Global Air Pollution and Climate Change Impacts: Lessons Learned From a NASA-Supported Science-Stakeholder Collaborative. *GeoHealth*, *4*(Issue 7). https://doi.org/10.1029/2020GH000270

ANMM, B. d. (2015). La contaminación del aire y los problemas respiratorios. *Revista de la Facultad de Medicina, 58*(5), 44-47. Obtenido de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0026-17422015000500044&lng=es&nrm=iso

Anyamba, A., Estes, J., Kline, K., & Collins, E. (2015). Remote Sensing. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition), Elsevier*, 419-424. https://doi.org/10.1016/B978-0-08-097086-8.72046-0.

Araki, S., Shima, M., & Yamamoto, K. (2018). Spatiotemporal land use random forest model for estimating metropolitan NO2 exposure in Japan. *Science of The Total Environment*, *634*, 1269-1277. https://doi.org/10.1016/j.scitotenv.2018.03.324

Ariana, C. (2021). Redes neuronales recurrentes: Análisis de los modelos especializados en datos secuenciales. *ECONSTOR*, (797).

Banerjee, C., Mukherjee, T., & Pasiliao, E. (2019). An Empirical Study on Generalizations of the ReLU Activation Function. *Proceedings of the 2019 ACM Southeast Conference*, 164–167. https://doi.org/10.1145/3299815.3314450

Bello, A. M., Cuta, J. A., & García, E. K. (2019). Técnicas de imputación para datos de precipitación máxima mensual en la zona central de Boyacá. *Rev. Revista Ingeniería, Investigación y Desarrollo*, *19*, 64-79.

Berrick, S. W., Leptoukh, G., Farley, J. D., & Rui, H. (2009). Giovanni: A Web Service Workflow-Based Data Visualization and Analysis System. *IEEE Transactions on Geoscience and Remote Sensing, 47*(1), 106-113. doi:10.1109/TGRS.2008.2003183

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, *549*(7671), Art. 7671. https://doi.org/10.1038/nature23474

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Biau, G. (2012). Analysis of a Random Forest Model. *Journal of Machine Learning Research*.

Breiman, L. (2001). Random Forest. *Kluwer Academic Publishers.*, (45), 5–32.

Brown, James Dean. “The coefficient of determination.” *Shiken: JALT Testing & Evaluation SIG Newsletter.*, vol. 7, no. 1, 2003, pp. 14-16.

Brunekreef, B., & Holgate, S. (2002). Contaminación atmosférica y salud. *The Lancet, 360*(9341), 1233-1242. doi:10.1016/S0140-6736(02)11274-8

Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., & Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy. *Atmospheric Environment*, *41*(14), 2967–2995. https://doi.org/10.1016/j.atmosenv.2006.12.013

Calopiña, D. C. (2014). *Diseño de una red neuronal artificial para la demanda de la predicción eléctrica*. Universidad Nacional de Loja.

Cardellino, F. (2021, March 22). *Tutorial para un clasificador basado en bosques aleatorios: cómo utilizar algoritmos basados en árboles para el aprendizaje automático*. freeCodeCamp. Retrieved October 11, 2022, from https://www.freecodecamp.org/espanol/news/random-forest-classifier-tutorial-how-to-use-tree-based-algorithms-for-machine-learning/

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, *91*(4), 045002. https://doi.org/10.1103/RevModPhys.91.045002

Carré, J., Gatimel, N., Moreau, J., Parinaud, J., & Léandri, R. (2017). Does air pollution play a role in infertility? a systematic review. *Environmental Health*, (82). https://doi.org/10.1186/s12940-017-0291-8.

Chai, T., & Draxler, R. R. (2014). Root means square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development*, *7*(3), 1525-1534. doi:10.5194/gmdd-7-1525-2014

Chardon Schirm, I. (2019). *PREDICCIÓN DE RENTAS EN SANTIAGO DE CHILE UTILIZANDO ALGORITMOS DE APRENDIZAJE AUTOMÁTICO*. Universidad de Chile.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U. A., Katsouyanni, K., Janssen, N. A. H., Martin, R. V., Samoli, E., Schwartz, P. E., Stafoggia, M., Bellander, T., Stark, M., Wolf, K., Vienneau, D., … Hoek, G. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment International*. https://doi.org/10.1016/j.envint.2019.104934

Coronel Carbo, J., & Marzo Páez, N. (2017). La promoción de salud para la creacion de entornos saludables en América Latina y el Caribe. *MEDISAN, 21*(12), 3414-3423. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017001200016&lng=es&tlng=es

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forest. In *Ensemble Machine Learning*. Zhang C. Ma. Y. https://doi.org/10.1007/978-1-4419-9326-7_5

Dahiya, S., Anhäuser, A., Farrow, A., Thieriot, H., Chanchal, A., & Myllyvirta, L. (2020). Ranking the World's Sulfur Dioxide (SO2) Hotspots: 2019-2020. *Delhi Center for Research on Energy and Clean Air-Greenpeace India: Chennai*, (48).

Data Products - Sentinel-5P Mission - Sentinel Online - Sentinel Online. (n.d.). Sentinel Online. Retrieved September 13, 2022, from https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-5p/data-products

EARTHDATA-NASA. (n.d.). EARTHDATA (Open Access for Open Science). Retrieved septiembre 13, 2022, from https://www.earthdata.nasa.gov/

ECMWF, C. E. (30 de 06 de 2022). *Datos horarios de ERA 5 sobre niveles de presion desde 1979 hasta el presente* . Obtenido de https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview

Espín Alarcón, G. E. (2011). *Contaminación atmosférica por el polvo PM-10, producto de la explotación minera y la regeneraación urbano-industrial de la ciudad de Guayaquil.* Universidad de Guayaquil, Guayaquil. Obtenido de http://repositorio.ug.edu.ec/handle/redug/506

Esri. (2019). *El tamaño de la celda y el remuestreo en análisis*. https://desktop.arcgis.com/es/arcmap/10.7/extensions/spatial-analyst/performing-analysis/cell-size-and-resampling-in-analysis.htm

European Center for Medium-Range Weather Forecasts (ECMWF). (n.d.). *ERA5: data documentation*. Retrieved September 12, 2022, from https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Howtoacknowledge,citeandrefertoERA5

Europea, A. E. (29 de 06 de 2022). *Centinela-5P*. Obtenido de https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-5p

Farrow, A., Miller, K. A., & Myllyvirta, L. (2020). *TOXIC AIR: THE PRICE OF FOSSIL FUELS.* https://www.greenpeace.org/usa/wpcontent/uploads/2020/02/The-Price-of-Fossil-Fuels-full-report.pdf)

F. Bornman, J., W. Barnes, P., A. Robinson, S., L. Ballaré, C., D. Flint, S., & M. Caldwell, M. (2015). Solar ultraviolet radiation and ozone depletion-driven climate change: Effects on terrestrial ecosystems. *Photochemical & Photobiological Sciences*, *14*(1), 88–107. https://doi.org/10.1039/C4PP90034K

Galvan Sala, D. A. (2021). *Comparativa de técnicas para la prevención del sobreajuste en redes neuronales*. Trabajo de fin de grado de la Universidad de Alicante.

García, M., Ramírez, H., Fuentes, M., Arias, S., & Espinosa, M. (2014). Comportamiento de los vientos dominantes y su influencia en la contaminación atmosférica en la zona metropolitana de Guadalajara, Jalisco, México. *Revista Iberoamericano de Ciencias*, *1*(2), 97-116.

Gavilánez Barrionueevo, C. S. (2021). *Estimación de la calidad del aire en ambientes interiores en laboratorios químicos de la Universidad Técnica de Ambato utilizando redes neuronales artificiales.* Universidad Técnica de Ambato, Ambato. Obtenido de http://repositorio.uta.edu.ec/handle/123456789/32088

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Mach Learn*, (63), 3-42. DOI 10.1007/s10994-006-6226-1

Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*, *21*, 65. https://doi.org/10.4103/1735-1995.189646

Giovanni NASA. (n.d.). *Estimación de las precipitaciones de satélite de TRMM y GPM.* Retrieved September 13, 2022, from https://giovanni.gsfc.nasa. gov/giovanni/

Global Modeling and Assimilation Office (GMAO). (2021). *MERRA-2 tavgM_2d_ocn_Nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation,Ocean Surface Diagnostics V5.12.4*. Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Retrieved septiembre 13, 2022, from DOI: 10.5067/4IASLIDL8EEC

González Duque, R. (2011). *Python para todos*. http://mundogeek.net/tutorial-python/

Google Developers (s.f.). (n.d.). *Sentinel-5P Data Catalog*. Retrieved September 12, 2022, from https://developers.google.com/earthengine/datasets/catalog/sentinel-5p

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*(ISSN 0034-4257), 18-27. https://doi.org/10.1016/j.rse.2017.06.031.

Guo, C., Liu, G., & Chen, C.-H. (2020). Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network. *Wireless Communications and Mobile Computing*. doi: https://doi.org/10.1155/2020/8854649

Graves, A. (2013). *Generating Sequences With Recurrent Neural Networks*. https://doi.org/10.48550/ARXIV.1308.0850

Guan, W.-J., Zheng, X.-Y., Chung, K. F., & Zhong, N.-S. (2016). Impact of air pollution on the burden of chronic respiratory diseases in China: Time for urgent action. *The Lancet*, *388*(10054), 1939–1951. https://doi.org/10.1016/S0140-6736(16)31597-5

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). STATUS OF AUTOMATIC CALIBRATION FOR HYDROLOGIC MODELS: COMPARISON WITH

MULTILEVEL EXPERT CALIBRATION. *Journal of Hydrologic Engineering*, *4*(135-143). DOI: 10.1061/(ASCE)1084-0699

Han, L., Zhou, W., Li, W., Meshesha, D. T., Li, L., & Zheng, M. (2015). Meteorological and urban landscape factors on severe air pollution in Beijing. *Journal of the Air & Waste Management Association*, *65*(7), 782–787. https://doi.org/10.1080/10962247.2015.1007220

Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of *Scikit-learn* Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348–361. https://doi.org/10.3102/1076998619832248

Herrera Burgos , J. A. (2019). *Sistema basado en Redes Neuronales Artificiales orientado a la predicción del dioxido de carbono (CO2) como índice de contaminación, en la zona céntrica de la ciudad de Santo Domingo.* Universidad Regional Autónoma de los Andes "UNIANDES", Santo Domingo.

Hesterberg, T. W., Bunn, W. B., McClellan, R. O., Hamade, A. K., Long, C. M., & Valberg, P. A. (2009). Critical review of the human data on short-term nitrogen dioxide (NO2) exposures: Evidence for NO2 no-effect levels. *Critical Reviews in Toxicology*, *39*, 743-781. https://doi.org/10.3109/10408440903294945

Hinton, G., Srivastava, N., & Swersky, K. (2012). Overview of mini-batch gradient descent.

Hosseini, V., & Shahbazi, H. (2016). Urban Air Pollution in Iran. *Iranian Studies*, *49*(6), 1029–1046. https://doi.org/10.1080/00210862.2016.1241587

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., & Liu, Y. (2018). Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain. *Environmental Pollution*, *242*, 675-683. https://doi.org/10.1016/j.envpol.2018.07.016

Hutton, G. (2011). Introduction. In Air Pollution: Global Damage Costs of Air Pollution from 1900 to 2050. *Copenhagen Consensus Center*, 4-6. http://www.jstor.org/stable/resrep16318.5)

Jacinto Herrera, R. (2019). *Redes Neuronales para la predicción de contaminación del aire en Carabayllo-Lima.* Universidad Nacional Federico Villareal, Lima.

Jacob, J., & NASA/GSFC/HSL. (2021). *FLDAS Noah Land Surface Model L4 Central Asia Daily 0.01 x 0.01 degree*. Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Retrieved septiembre 13, 2022, from DOI: 10.5067/VQ4CD3Y9YC0R

Jerrett, M. (2015). The death toll from air-pollution sources. *Nature*, *525*(7569), Art. 7569. https://doi.org/10.1038/525330a

Jin, J. (2020). *Desarrollo de una Herramienta de Integración de datos de Imágenes de Satélite en Google Earth Engine*. Tesis de Ingeniería Informática de la Universidad Politécnica de Madrid. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://oa.upm.es/63348/1/TFG_JIA HAO_JI.pdf

Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., & Xie, G. (2018). Air Quality Prediction: Big Data and Machine Learning Approaches. *International Journal of Environmental Science and Development*, *9*(1). DOI: 10.18178/ijesd.2018.9.1.1066

Katulski, R. J., Namieśnik, J., Sadowski, J., Stefański, J., & Wardencki, W. (2011). Monitoring of Gaseous Air Pollution. In *The impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*. Mohamed K. Khallaf. DOI: 10.5772/20771

Kigman, D. P., & Lei Ba, J. (2017). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*.

Kumar, B. (2020, November 17). *Python Epoch to Date Time*. Python Guides. Retrieved September 30, 2022, from https://pythonguides.com/python-epoch-to-datetime/

Kort, E., Frankenberg, C., Miller , C., & Oda, T. (2012). Space-based observations of megacity carbon dioxide. *Geophysical Research Letters*, 39 (17). DOI: 10.1029/2012GL05

Lacasaña Navarro, M., Aguilar Garduño, C., & Romieu, I. (1999). Evolución de la contaminación del aire e impacto de los programas de control en tre megaciudades ded América Latina . *Salud Pública de México*, 41, 203-215.

Lama , S., Houweling, S., Boersma, K., Aben, I., van der Gon, H., Dolman, A., . . . Lorente, A. (2019). Quantifying burning efficiency in Megacities using NO2/CO ratio from the Tropospheric Monitoring Instrument (TROPOMI). *Atmospheric Chemistry and Physics. Discuss, 1*. DOI:10.5194/acp-2019-1112

Llanque Chana, J. (2003). *Efectos de la Contaminación Atmosférica en el clima Urbano y Calidad Ambiental de Arquipa.* Universidad Nacional San Agustín Arequipa, Perú.

Lavín Pallero, E. (2021). *Funciones de pérdida dinámicas en Machine Learning: Aplicación del Error Cuadrático Medio Dinámico en Redes Neuronales Artificiales*. Trabajo Fin de Grado de Ingeniería en Tecnologías Industriales en la Universidad Politécnica de Madrid.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, *525*(7569), 367–371. https://doi.org/10.1038/nature15371

Le, V.D., & Cha, S. K. (2018). Real-time Air Pollution prediction model based on Spatiotemporal Big data. *Computers and Society*. https://doi.org/10.48550/arXiv.1805.00432 Focus to learn more

Li, B., Beaudoing, H., & NASA/GSFC/HSL. (2019). GLDAS Catchment Land Surface Model L4 daily 0.25 x 0.25 degree GRACE-DA1 V2.2. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. DOI: 10.5067/TXBMLX370XX8

Li, A., Feng, M., Li, Y., & Liu, Z. (2016). Application of outlier mining in insider identification based on Boxplot method. *Procedia Computer Science*, *91*(ISSN 1877-0509), 245-251. https://doi.org/10.1016/j.procs.2016.07.069

Liu, B., Yan, S., Li, J., Li, Y., Lang, J., & Qu, G. (2021). A Spatiotemporal Recurrent Neural Network for Prediction of Atmospheric PM2.5: A Case Study of Beijing. IEEE Transactions on Computational Social Systems, 8(3), 578–588. https://doi.org/10.1109/TCSS.2021.3056410

Li, J., Shao, X., & Zhao, H. (2018). An Online Method Based on Random Forest for Air Pollutant Concentration Forecasting. 2018 37th Chinese Control Conference (CCC), 9641–9648. https://doi.org/10.23919/ChiCC.2018.8483621

Li, L., Qian, J., Ou, C.-Q., Zhou, Y.-X., Guo, C., & Guo, Y. (2014). Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. *Environmental Pollution*, *190*, 75–81. https://doi.org/10.1016/j.envpol.2014.03.020

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). *A Critical Review of Recurrent Neural Networks for Sequence Learning*. https://doi.org/10.48550/ARXIV.1506.00019

Li, R., Cui, L., Meng, Y., Zhao, Y., & Fu, H. (2019). Satellite-based prediction of daily SO2 exposure across China using a high-quality random forest-spatio temporal Kriging (RF-STK) model for health risk assessment. *Atmospheric Environment*, *208*, 10-19. https://doi.org/10.1016/j.atmosenv.2019.03.029

Llosa, Z. B. (2010). 50 Biocenosis • Vol. 23 (1) 2010 CONTAMINACIÓN ATMOSFÉRICA EN LA MESETA CENTRAL DE COSTA RICA.

Loenen, A. F., Huijnen, V., Douros, J., Miguens, L., & Biserkov, K. G. (2021). AIR-Portal: Service for Urban Air Quality Monitoring. In: Urbach, H.P., Yu, Q. (eds) 6th International Symposium of Space Optical Instruments and Applications. Space Technology Proceedings. *Springer, Cham*, *7*. https://doi.org/10.1007/978-3-030-56488-9_13

Manisalidis, L., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Front Public Health*. DOI: 10.3389/fpubh.2020.00014

Marlier, M. E., Jina, A. S., Kinney, P. L., & DeFries, R. S. (2016). Extreme Air Pollution in Global Megacities. *Current Climate Change Reports*, *2*(1), 15–27. https://doi.org/10.1007/s40641-016-0032-z

Martínez Ataz, E., & Díaz de Mera Morales, Y. (2004). *CONTAMINACIÓN ATMOSFÉRICA*. Ediciones de la Universidad de Castilla-La Mancha.

Martínez Mateo, E. (2021). *Contaminación atmosférica een la Península Ibérica medida con Sentinel 5P.* Universidad de Valencia.

Michél Cortés, R. E., Herbas Barrancoa, J. P., Paz Aldana, M., & Cortez Alemán, C. H. (2013). Análisis del grado de contaminación del aire en la ciudad de Tarija. *Revista Ventana Científica, 1*(5), 53-65.

Montero López , I. L., Vinueza Veloz, M. F., Castillo López, G. A., Ruano Ipiales, D. S., & Martín Barceló, N. (2020). Afecciones respiratorias y contaminación ambiental en Riobamba, Ecuador. *Correo Científico Médico (CCM), 24*(1). Obtenido de http://revcocmed.sld.cu/index.php/cocmed/article/view/3368

Morales Baquero, R., Martínez, C. P., Reche, I. (2001). Ecosistemas de alta montaña, las atalayas de la troposfera. *Revista de Ecología y medio ambiente*, (3).

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bigner, R. L., Harmel, R. D., & Veith, T. L. (2007). MODEL EVALUATION GUIDELINES FOR SYSTEMATIC QUANTIFICATION OF ACCURACY IN WATERSHED SIMULATIONS. *Academia Accelerating the world's research.*, 16.

Moscoso Vanegas, D., Astudillo Alemán , A., & Morales Pérez, M. (2018). Inventario de emisiones atmosféricas provenientes de fuentes fijas de combustión del parque industrial del cantón Cuenca-Ecuador. *Revista Centro Azúcar, 45*, 2223-4861.

Moyano Tobar, C. M. (2017). *Estimación de la contaminación del aire generada por efecto de la circulación vehicular motorizada en la Av. 10 de agosto de la ciudad de Cuenca - Ecuador, usando la herramienta de microsimulación de tránsito Aimsun 8.1.* Trabajo de titulación de maestría en tránsito, transporte y seguridad vial.

Muñoz D., A. M., Paz V., J. J., Quiroz P., C. M. (2007). Efectos de la contaminación atmosférica sobre la salud de adultos que laboran en diferentes niveles de exposición. *Rev. Fac. Nac. Salud Pública*, *25*(2), 85-94.

Mutanga, O., & Kumar, L. (2019). Google Earth Engine Applications. *Remote Sensing*, *11*(5). https://doi.org/10.3390/rs11050591

Natal Pérez, S. (2021). INFLUENCIA DE LA METEOROLOGÍA Y EL TRÁFICO EN LA CALIDAD DEL AIRE DE LA CIUDAD DE MADRID. *Universidad Internacional Menéndez Pelayo*.

Niño Laguado, J. M. (2018). *PROPUESTA PARA LA IMPLEMENTACIÓN DE UN MÉTODO DE PRONÓSTICO DE LA DEMANDA DEL MATERIAL DE EMBALAJE EN UNA EMPRESA DISTRIBUIDORA DE DISPOSITIVOS MÉDICOS*.

NLDAS project. (2021). *NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V2.0 (NASA/GSFC/HSL)*. Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Retrieved septiembre 13, 2022, from DOI: 10.5067/T4OW83T8EXDO

Ortiz Ramírez, A. (2010). *Python como primer lenguaje de programación*. Departamento de Tecnologías de Información y Computación.

Páez Freire, C. F. (2022). *Análisis de la variación de la calidad del aire debido al confinamiento por COVID-19 en ciudades de América Latina mediante imágenes satelitales Sentinel-5P*. Tesis de Ingeniería Geográfica y del Medio Ambiente de la Universidad de las Fuerzas Armadas. http://repositorio.espe.edu.ec/bitstream/21000/28761/1/T-ESPE-050983.pdf

Palacios Espinoza, E., & Espinoza Molina, C. (2014). Contaminación del aire exterior. Cuenca-Ecuador, 2009-2013. Posibles Efectos en la salud. *Revista de la Facultad de Ciencias Médicas de la Universidad de Cuenca, 32*(2), 6-17. Obtenido de https://publicaciones.ucuenca.edu.ec/ojs/index.php/medicina/article/view/883

Park, H., Jeong , S., Park , H., Labzovskii, L., & Bowman, K. (2021). An assessment of emission characteristics of Northern Hemisphere cities using spaceborne observations of CO2, CO, and NO2. *Remote Sensing of Enviroment*, 254, 112246. doi:10.1016/j.rse.2020.112246

Pedraza Camelo, J. C. (2019). *Prototipo de un modelo de machine learning para la predicción de partículas de contaminación atmosférica finas en la localidad de Kennedy en Bogotá.* Universidad Distrital Francisco José de Caldas. Obtenido de https://repository.udistrital.edu.co/handle/11349/15772?show=full

Peña Murillo, S. E. (2018). Impacto de la contaminación atmosférica en dos principales ciudades del Ecuador. *Revisra Universidad y Sociedad, 10*(2), 289-293.

Pérez Ortiz, J. A. (2002). *Modelos Predictivos basados en Redes Neuronales Recurrentes en tiempo discreto*. Universidad de Alicante.

Perilla, G. A., & François, J. (2020). Google Earth Engine (GEE): una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube. *Investigaciones Geográficas*, (UNAM eISSN: 2448-7279). https://doi.org/10.14350/rig.59929

Phuleria, H. (2013). The role of air quality monitoring networks in supporting health research and legislation. *BMC Proceedings*, (7 (Suppl 5:O11). https://doi.org/10.1186/1753-6561-7-S5-O11

Puri, P., Nandar, S. K., Kathuria, S., & Ramesh, V. (2017). Effects of air pollution on the skin: A review. *Indian Journal of Dermatology, Venereology and Leprology*, *83*, 415. https://doi.org/10.4103/0378-6323.199579

Qi, Z., Wang, T., Song, G., Hu, W., & Zhang, Z. (2018). Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-Grained Air Quality. *IEEE Transactions on Knowledge and Data Engineering*, *30*(12), 2285-2297. Doi: 10.1109/TKDE.2018.2823740

Ramiréz Gutiérrez, M. Á. (2017). Uso de modelos Lineales Generalizados (MLG) para la interpolación espacial de PM10 utilizando imágenes satelitales Landsat para la ciudad de Bogotá, Colombia . *Perspectiva Geografica*, 22(2), 105-121.

Rani Hemamalini, R., Vinodhini, R., Shanthini, B., Partheeban, P., Charumathy, M., & Cornelius, K. (2022). Air quality monitoring and forecasting using smart drones and recurrent neural network for sustainable development in Chennai city. *Sustainable Cities and Society*, *85*, 104077. https://doi.org/10.1016/j.scs.2022.104077

Robinson, D. L. (2005). Air pollution in Australia: Review of costs, sources and potential solutions. *Health Promotion Journal of Australia*, *16*(3), 213–220. https://doi.org/10.1071/HE05213

Rodríguez Sánchez, A., Salmerón Gómez, R., & García, C. (2022). The coefficient of determination in the ridge regression. *Communications in Statistics - Simulation and Computation*, *51*(1), 201–219. https://doi.org/10.1080/03610918.2019.1649421

Rohde, R. A., & Muller, R. A. (2015). Air Pollution in China: Mapping of Concentrations and Sources. *PLOS ONE*, *10*(8), e0135749. https://doi.org/10.1371/journal.pone.0135749

Romero, H., Irarrázaval, F., Opazo, D., Salgado, M., & Smith, P. (2010). Climas urbanos y contaminación atmosférica en Santiago de Chile. *EURE (Santiago)*, *36*(109), 35-62. http://dx.doi.org/10.4067/S0250-71612010000300002

Romero Placeres, M., Diego Olite, F., & Álavrez Toste, M. (2006). La contaminación del aire: su repercusión como problemas de salud. *Rev Cubana Hig Epidemiol*, 44(2). Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-30032006000200008&lng=es.

Romero Saldaña, M. (2016). Pruebas de bondad de ajuste a una distribución normal. *Rev. Enfermería del Trabajo*, *6*(ISSN-e 2174-2510), 114.

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences*, *8*(12), Art. 12. https://doi.org/10.3390/app8122570

Sáenz García, G. (2016). *Estudio y caracterización de la tropopausa: térmica, dinámica y química.* Universidad de Estremadura.

Salazar, C. P., Castillo, G., & del Castillo, S. (2018). *Fundamentos básicos de Estadística*.

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2018). *Recent Advances in Recurrent Neural Networks*. https://doi.org/10.48550/ARXIV.1801.01078

Sanjeev, D. (2021). Implementation of Machine Learning Algorithms for Analysis and Prediction of Air Quality. *International Journal of Engineering Research & Technology (IJERT*, *10*.

Sanjuán de Caso, M. (2020). *Predicción de la calidad del aire de la ciudad de Madrid mediante técnicas de machine-learning.* Universitat Oberta de Catalunya (UOC). Obtenido de http://hdl.handle.net/10609/120066

Scornet, E., Biau, G., & Vert, J.-P. (2015). Consistency of random forests. *The Annals of Statistics*, *43*(4), 1716–1741. https://doi.org/10.1214/15-AOS1321

Sethi, J. K., & Mittal, M. (2019). Ambient Air Quality Estimation using Supervised Learning Techniques. *EAI Endorsed Transactions on Scalable Information Systems*, *6*(22). doi: 10.4108/eai.13-7-2018.159406

Silibello, C., Carlino, G., Stafoggia, M., Gariazzo, C., Finardi, S., Pepe, N., Radice, P., Forastiere, F., & Viegi, G. (2021). Spatial-temporal prediction of ambient nitrogen dioxide and ozone levels over Italy using a Random Forest model for population exposure assessment. *Air Quality, Atmosphere & Health*, *14*, 817-829. https://doi.org/10.1007/s11869-021-00981-4

Singh, S., Yadav, A., & Kumar, A. (2021). Prediction of Air Pollution Using Random Forest. *25*(4), pp. 19314-19322.

Siwek, K., & Osowski, S. (2016). DATA MINING METHODS FOR PREDICTION OF AIR POLLUTION. *Int. J. Appl. Math. Comput. Sci.*, *26*(2), 467-478. DOI: 10.1515/amcs-2016-0033

Ssenyunzi, R. c., Oruru, B., D'ujanga, F. M., Realini, E., Barindelli, S., Tagliaferro, G., Engeln, A. v., & Guiesen, N. v. d. (2020). Performance of ERA5 data in retrieving Precipitable Water Vapour over East African tropical region. *Advances in Space Research*, *65*(8), 1877-1893. https://doi.org/10.1016/j.asr.2020.02.003.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., de' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., & Schwartz, J. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment International*, *124*, 170-179. https://doi.org/10.1016/j.envint.2019.01.016

Stathakis, D. (2009). How many hidden layers and nodes? *International Journal of Remote Sensing*, *30*(8), 2133–2147. https://doi.org/10.1080/01431160802549278

10 introducción a las redes neuronales artificiales | Introducción al software estadístico R. (n.d.). Dae-Jin Lee. Retrieved October 11, 2022, from https://idaejin.github.io/courses/R/2019/euskaltel/introduccion-a-las-redes-neuronales-artificiales.html

Uzair, M., & Jamil, N. (2020). Effects of Hidden Layers on the Efficiency of Neural networks. *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 1–6. https://doi.org/10.1109/INMIC50486.2020.9318195

Vani, S., & Rao, T. V. M. (2019). An Experimental Approach towards the Performance Assessment of Various Optimizers on Convolutional Neural Network. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 331–336. https://doi.org/10.1109/ICOEI.2019.8862686

van Geffen, J. H. G. M., Eskes, H. J., Boersma, K. F., & Veefkind, J. P. (2021). *TROPOMI ATBD of the total and tropospheric NO2 data products.* https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBDNO2-data-products

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Senl, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., … Levelt, P. F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, *120*, 70-83. https://doi.org/10.1016/j.rse.2011.09.027.

Velumani, K., Lopez-Lozano, R., Madec, S., Guo, W., Gillet, J., Comar, A., & Baret, F. (2021). Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model: Impact of the Spatial Resolution. *Rev. Planet Phenomics*.

Venegas, L. E., & Mazzeo, N. A. (2012). *LA VELOCIDAD DEL VIENTO Y LA DISPERSIÓN DE CONTAMINANTES EN LA ATMÓSFERA*. II Congreso Latinoamericano de Ingeniería de Vientos.

Vu, B. N., Sánchez, O., Bi, J., Xiao, Q., Hansel, N. N., Checkley, W., Gonzales, G. F., Steenland, K., & Liu, Y. (2019). Developing an Advanced PM2.5 Exposure Model in Lima, Peru. Remote Sensing, 11(6), Art. 6. https://doi.org/10.3390/rs11060641

Xia, F., Xing, J., Xu, J., & Pan, X. (2022). The short-term impact of air pollution on medical expenditures: Evidence from Beijing. *Journal of Environmental Economics and Management*, *114*, 102680. https://doi.org/10.1016/j.jeem.2022.102680

Yang, J., Ji, Z., Kang, S., Zhang, Q., Chen, X., & Lee, S.-Y. (2019). Spatiotemporal variations of air pollutants in western China and their relationship to meteorological factors and emission sources. *Environmental Pollution*, *254*, 112952. https://doi.org/10.1016/j.envpol.2019.07.120

Zambrano Mera, J. U. (2014). *IDENTIFICACIÓN DE LA CONTAMINACIÓN ATMOSFÉRICA GENERADA POR BUQUES DE CARGA EN EL PUERTO DE GUAYAQUIL*. Tesis de grado de maestría en Ciencias Ambientales de la Escuela Politécnica del Litoral.

Zhan, Y., Luo, Y., Deng, X., Zhang, K., Zhang, M., Grieneisen, M. L., & Di, B. (2018). Satellite-Based Estimates of Daily NO2 Exposure in China Using Hybrid Random Forest and Spatiotemporal Kriging Model. *Envrion. Sci. Technol.*, *52*(7), 4180 - 4189. https://doi.org/10.1021/acs.est.7b05669

Zheng, Z., Yang, Z., Wu, Z., & Marinello, F. (2019). Spatial Variation of NO2 and Its Impact Factors in China: An Application of Sentinel-5P Products. *Remote Sensing*, *11*(16). https://doi.org/10.3390/rs11161939

11. ANEXOS

A. Anexo: Scripts Utilizados para la obtención de Imágenes Satelitales.

```
//Datos Sentinel 5P para NO2
var  N02  =  ee.ImageCollection ( 'COPERNICUS/S5P/OFFL/L3_NO2' )
  . fecha del filtro ( '2020-12-31' , ' 2021-01-01' ) ; //Selección de periodo temporal

//Datos para columna NO2 Troposferico
var  SentinelNO2Tropo  =  N02
  . seleccione ( 'tropospheric_NO2_column_number_density' )
  . filterBounds  ( geometría ) ;
var  NO2TropoData  =  ee.Image ( SentinelNO2Tropo . mediana ( )) ;
var  NO2TropoClip  =  NO2TropoData . clip  ( geometría ) ;
Mapa.addLayer  ( NO2TropoClip ,  {
  mín :  -0.0005375 ,
  máx :  0.0192044 ,
  palette: ['black', 'blue', 'purple', 'cyan', 'green', 'yellow', 'red']});
Export.image.toDrive({
  image: NO2TropoClip.select("tropospheric_NO2_column_number_density"),
  description: 'NO2_Troposferico',
  scale: 1100,
  region: geometry});
```

Figura A-1. Script para la obtención de imágenes satelitales Sentinel 5P para dióxido de nitrógeno NO_2.

```
//Datos Sentinel 5P para SO2
var S02 = ee.ImageCollection('COPERNICUS/S5P/OFFL/L3_SO2')
  .filterDate('2020-12-31', '2021-01-01'); //Selección de periodo temporal

//Datos para columna SO2 Total
var SentinelSO2Total = S02
  .select('SO2_column_number_density')
  .filterBounds (geometry);
var SO2TotalData = ee.Image(SentinelSO2Total.median());
var SO2TotalClip = SO2TotalData.clip (geometry);
Map.addLayer (SO2TotalClip, {
  max: 0.0002,
  min: 0.0,
  palette: ["black", "blue", "purple", "cyan", "green", "yellow", "red"]},
  'SO2 Total');
Export.image.toDrive({
  image: SO2TotalClip.select("SO2_column_number_density"),
  description: 'SO2_Total',
  scale: 1100,
  region: geometry});
```

Figura A-2. Script para la obtención de imágenes satelitales Sentinel 5P para dióxido de nitrógeno SO_2.

B. Anexo: Relación de variables de concentración con variables meteorológicas de NO_2 en Python.

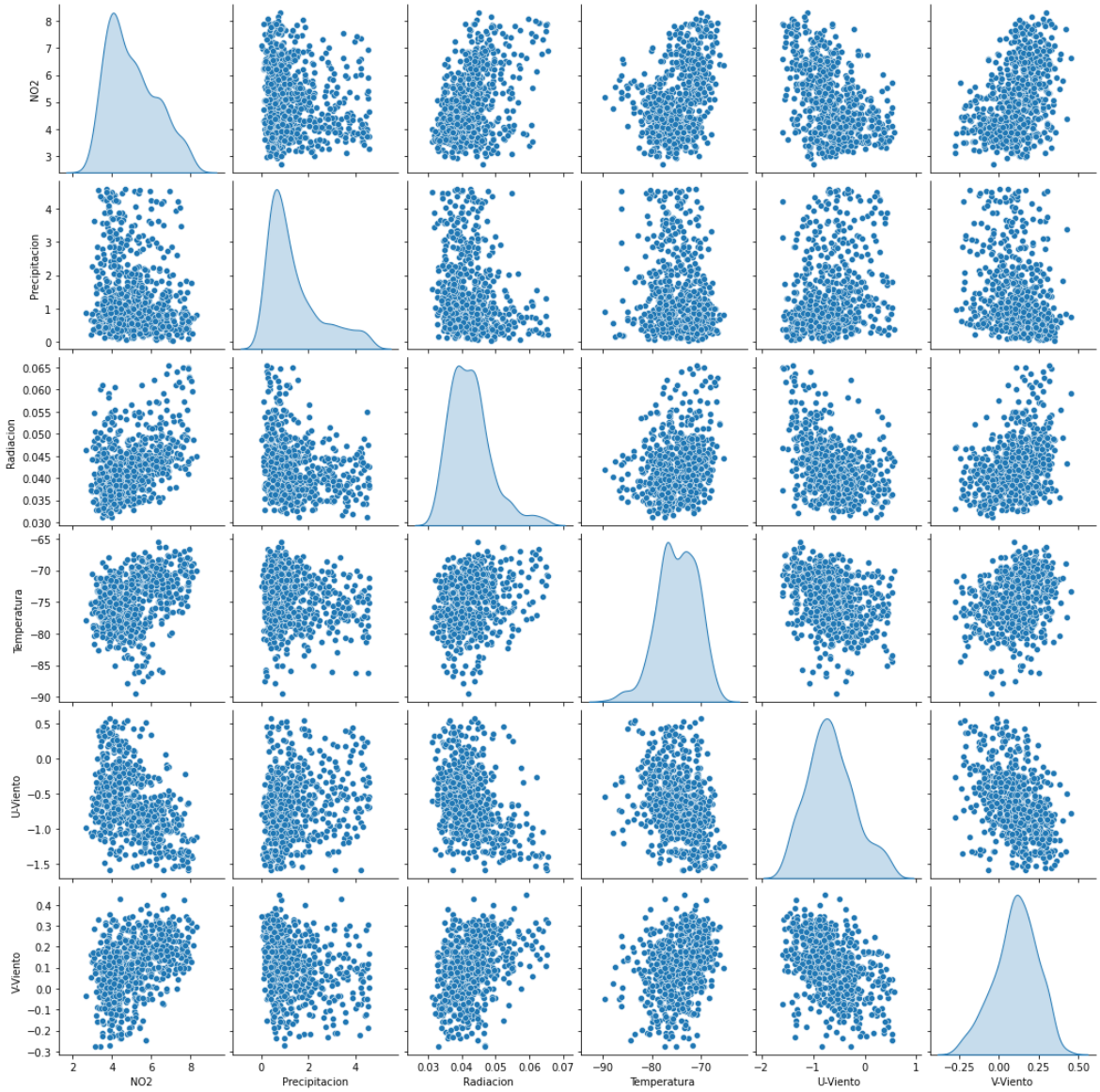

Figura B-1. Relación entre variables de concentración y variables meteorológicas de Cuenca para NO_2.

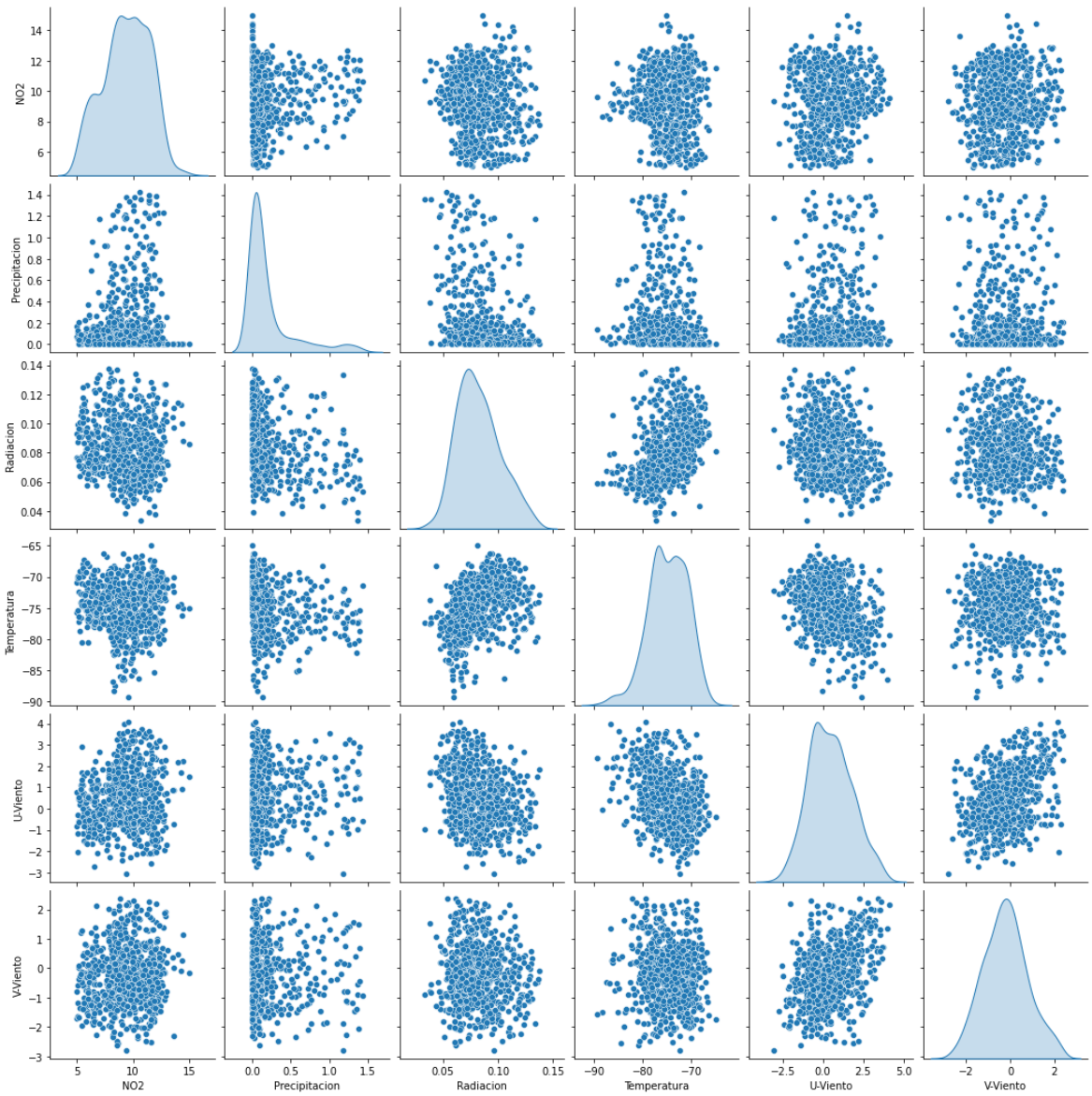

Figura B-2. Relación entre variables de concentración y variables meteorológicas de Guayaquil para NO_2.

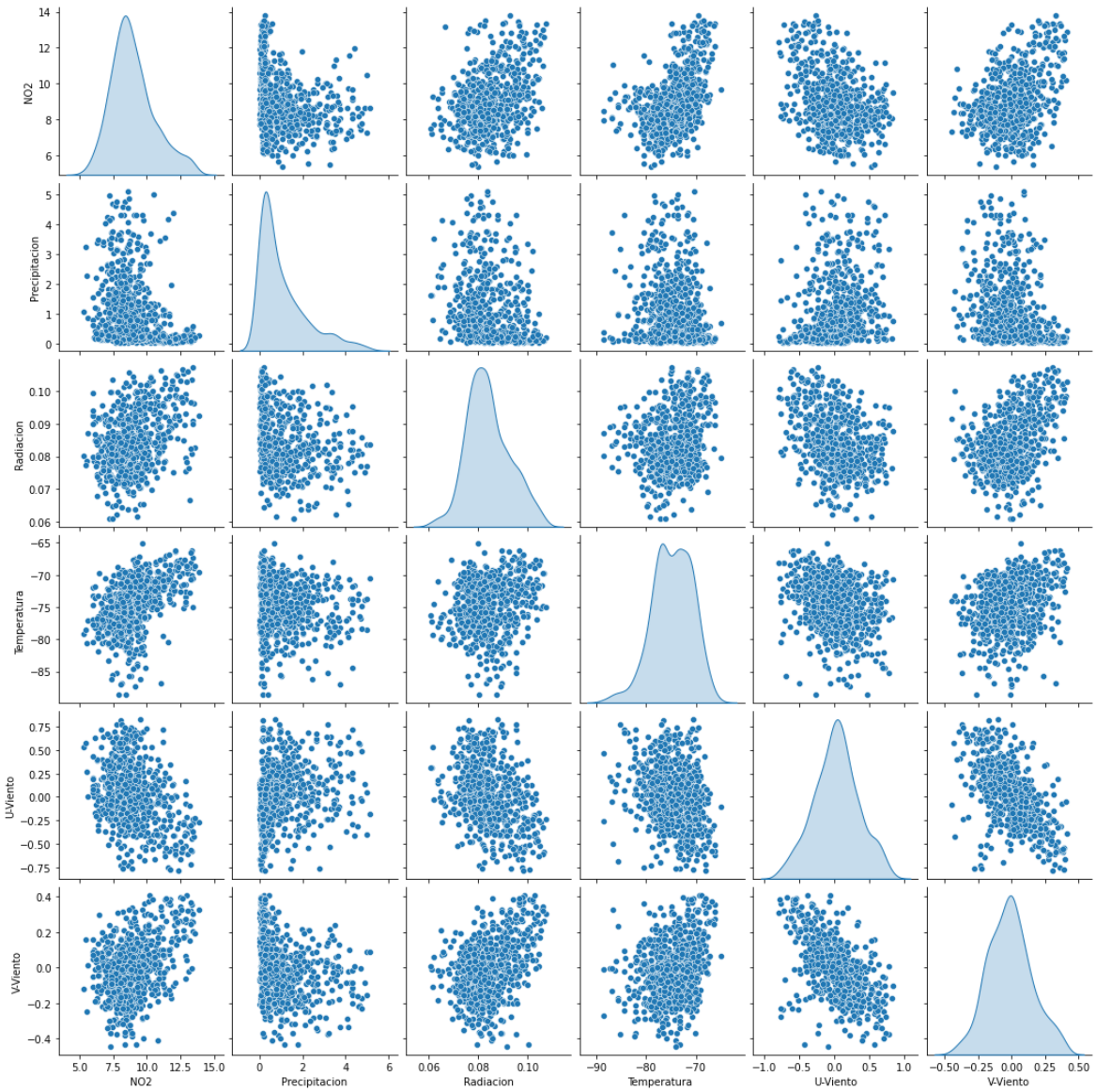

Figura B-3. Relación entre variables de concentración y variables meteorológicas de Quito para NO_2.

C. Anexo: Relación de variables de concentración con variables meteorológicas de SO_2 en Python.

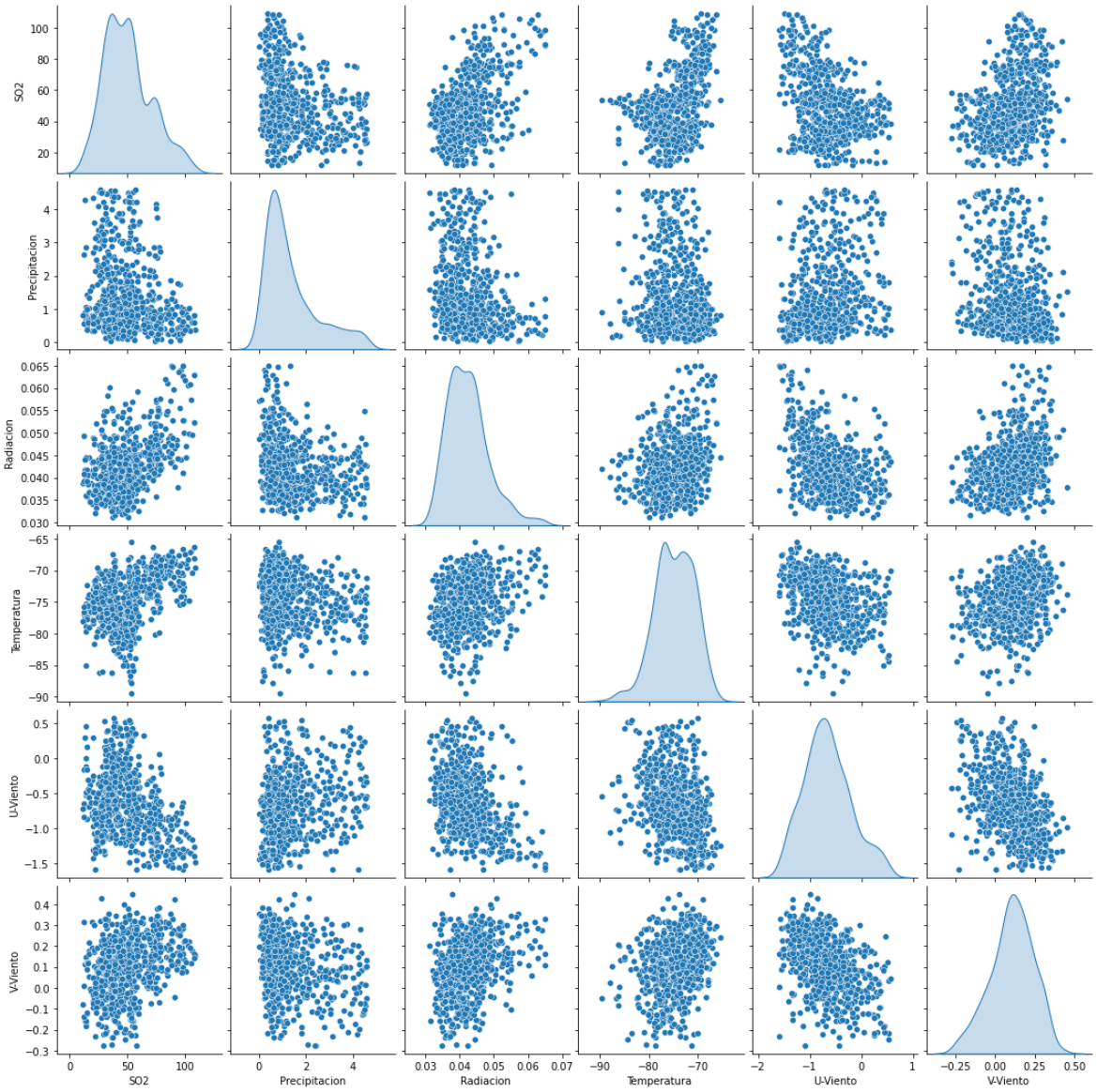

Figura C-1. Relación entre variables de concentración y variables meteorológicas de Cuenca para SO_2.

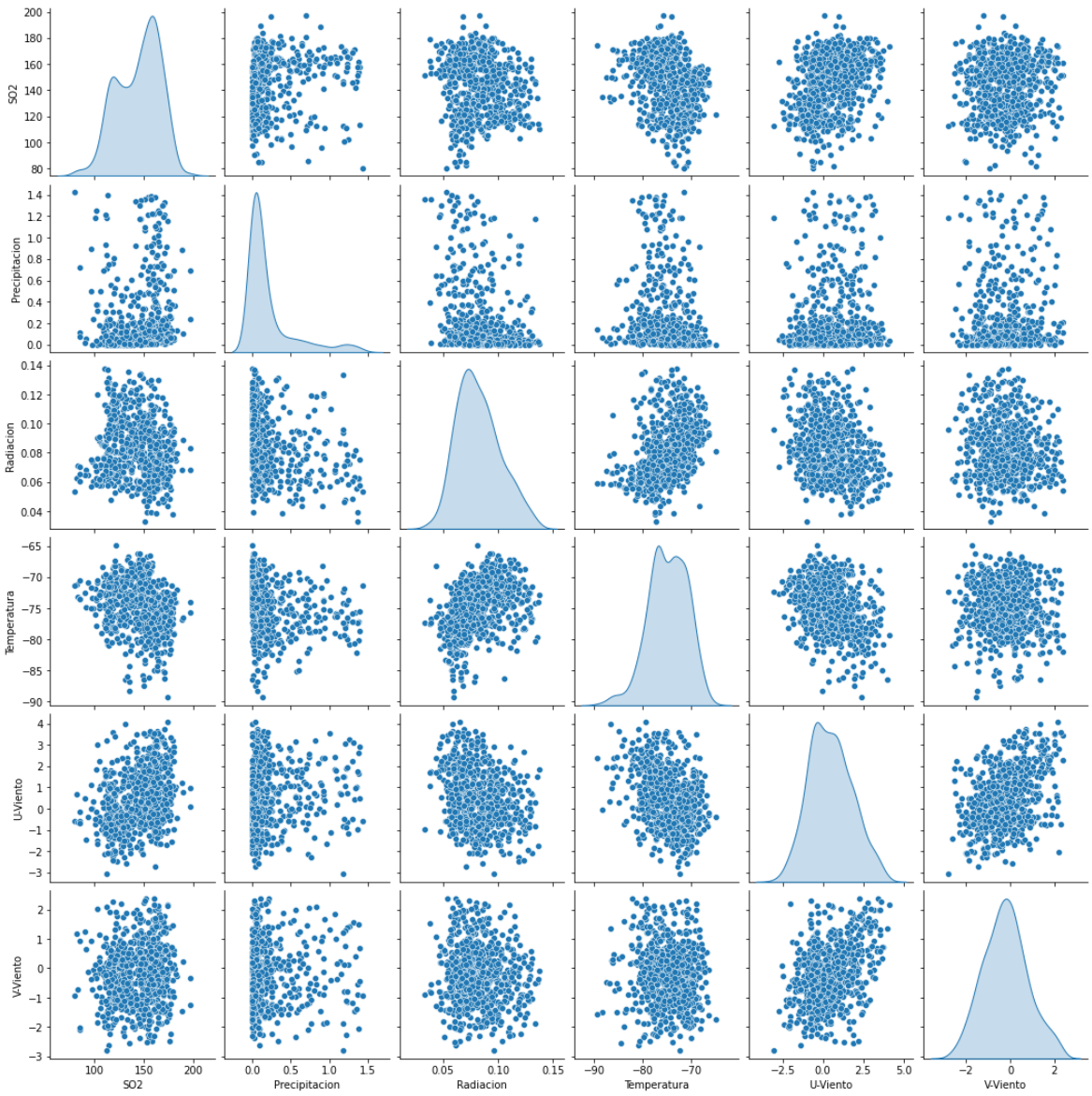

Figura C-2. Relación entre variables de concentración y variables meteorológicas de Guayaquil para SO_2.

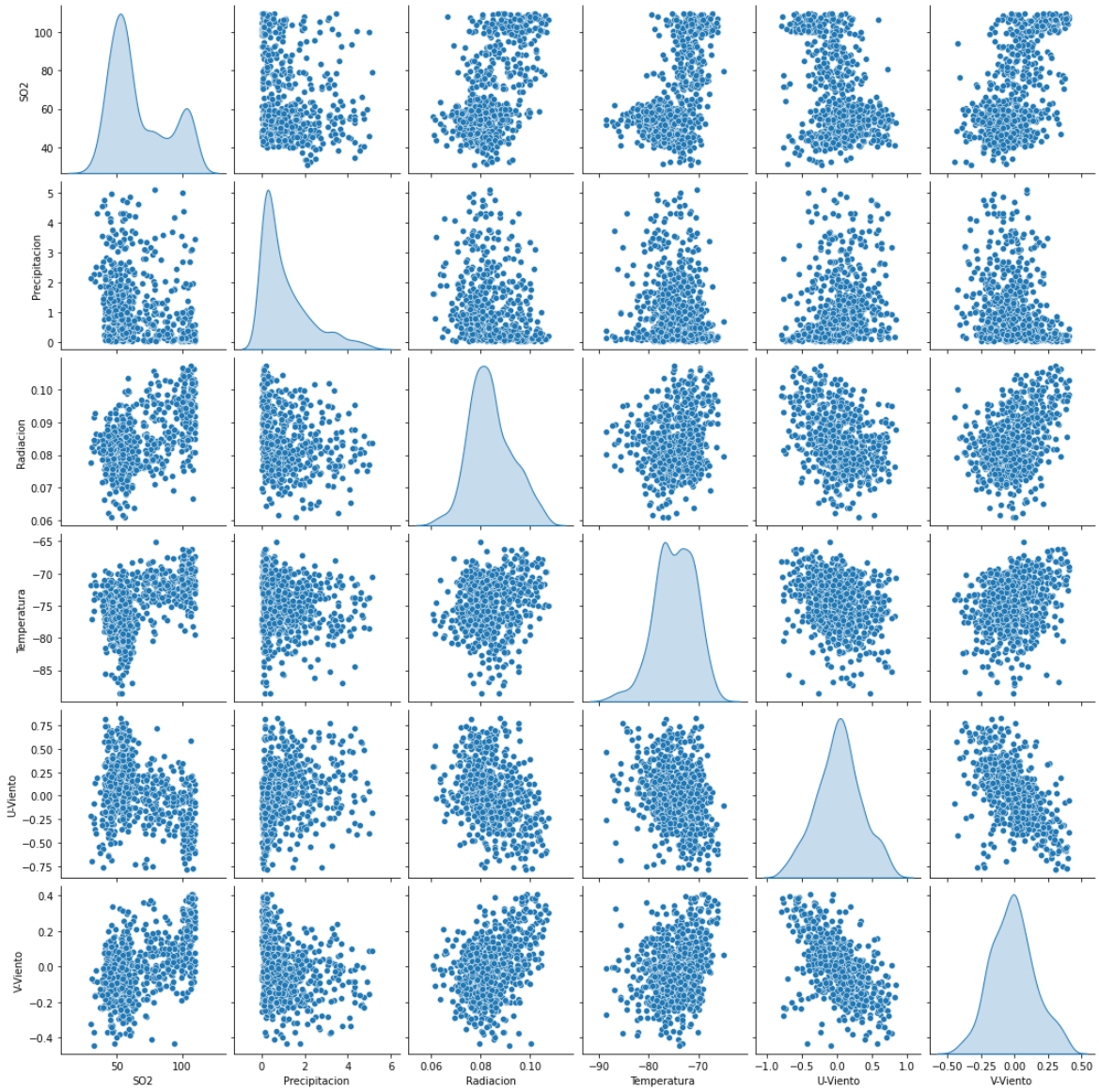

Figura C-2. Relación entre variables de concentración y variables meteorológicas de Quito para SO_2.

D. Anexo: Gráficas de dispersión de datos (NO_2) mediante Redes Neuronales Recurrentes con diferentes retrasos para 5 días.

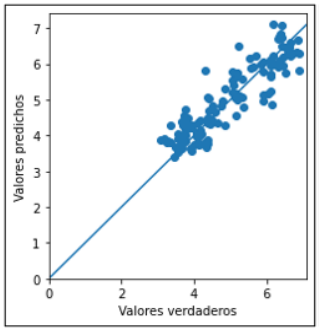

Gráfico D-1. Día 1 - retraso 0.

Gráfico D-2. Día 1 - retraso 1.

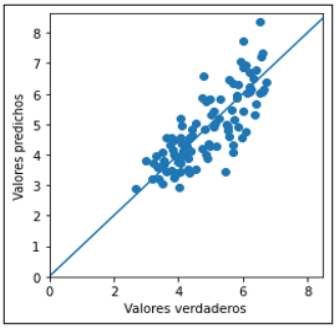

Gráfico D-3. Día 1 - retraso 2.

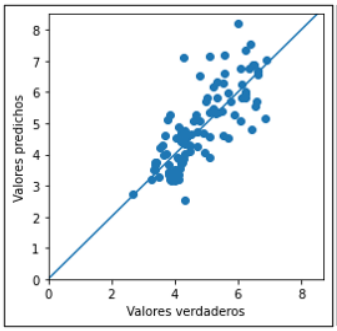

Gráfico D-4. Día 1 - retraso 3.

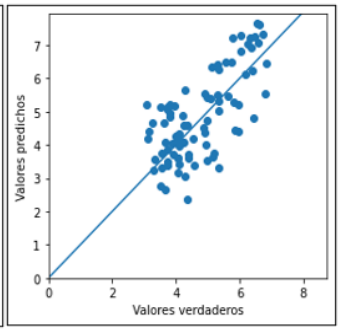

Gráfico D-5. Día 1 - retraso 4.

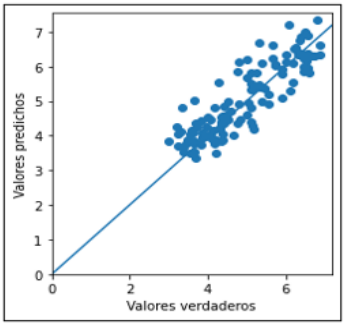

Gráfico D-11. Día 2 - retraso 0.

Gráfico D-12. Día 2 - retraso 1.

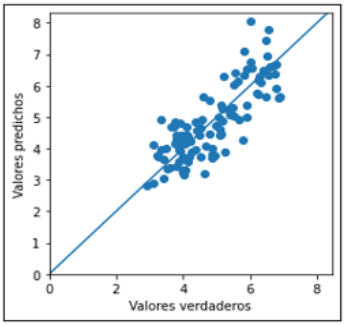

Gráfico D-13. Día 2 - retraso 2.

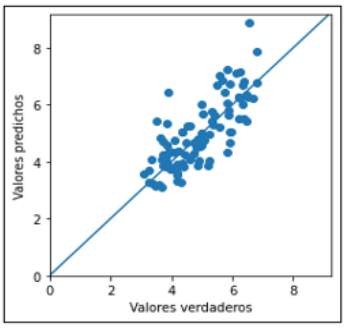

Gráfico D-14. Día 2 - retraso 3.

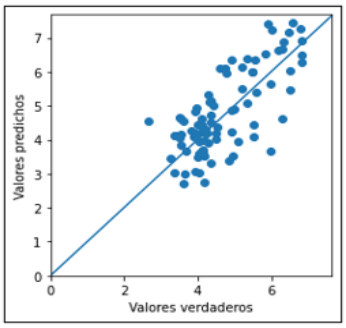

Gráfico D-15. Día 2 - retraso 4.

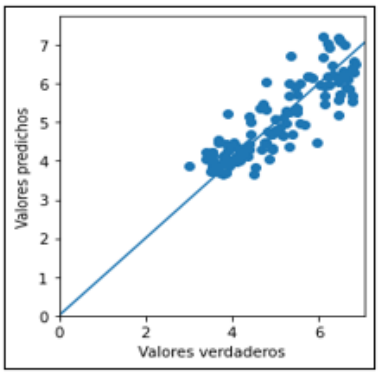

Gráfico D-21. Día 3 - retraso 0.

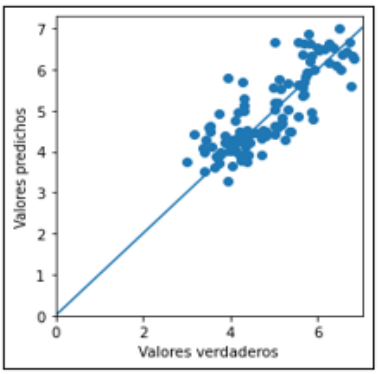

Gráfico D-22. Día 3 - retraso 1.

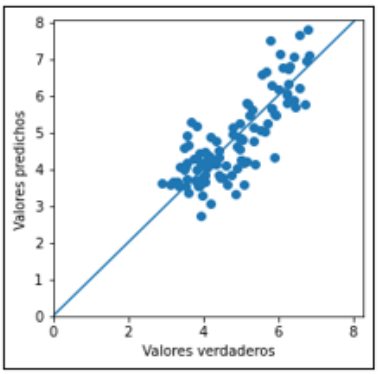

Gráfico D-23. Día 3 - retraso 2.

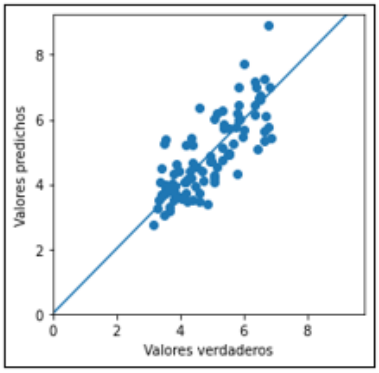

Gráfico D-24. Día 3 - retraso 3.

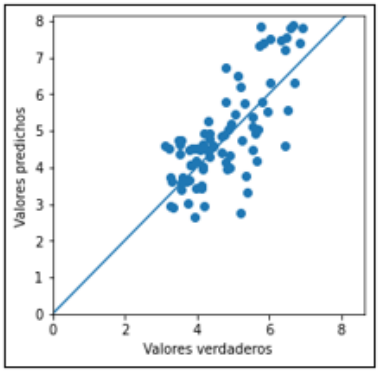

Gráfico D-25. Día 3 - retraso 4.

Gráfico D-31. Día 4 - retraso 0.

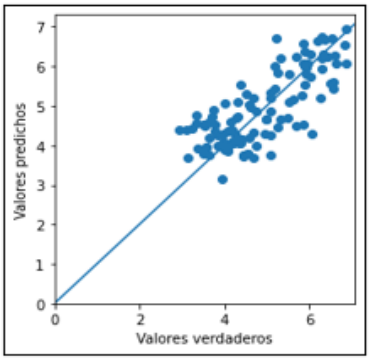

Gráfico D-32. Día 4 - retraso 1.

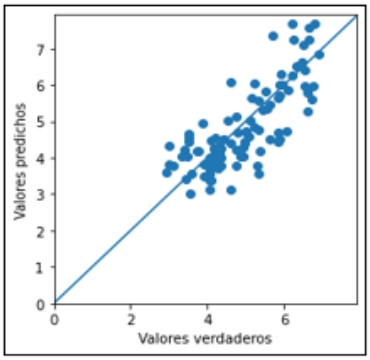

Gráfico D-33. Día 4 - retraso 2.

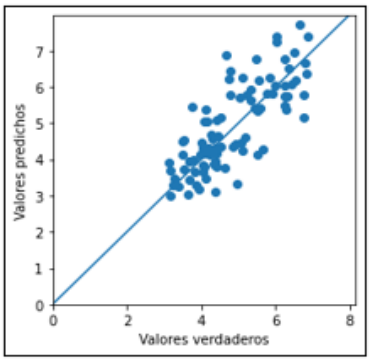

Gráfico D-34. Día 4 - retraso 3.

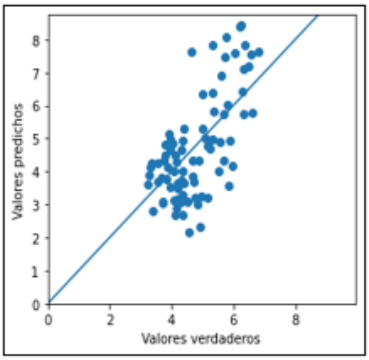

Gráfico D-35. Día 4 - retraso 4.

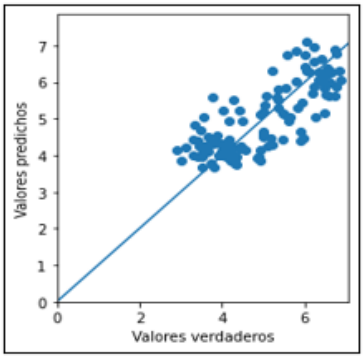

Gráfico D-41. Día 5 - retraso 0.

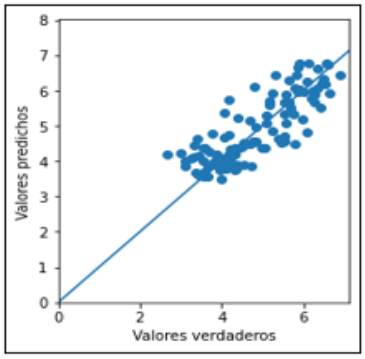

Gráfico D-42. Día 5 - retraso 1.

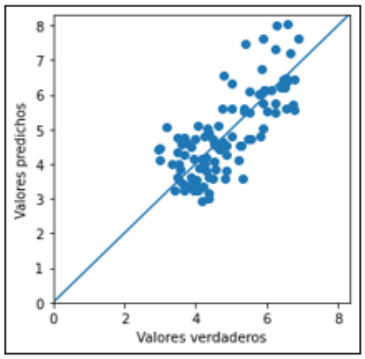

Gráfico D-43. Día 5 - retraso 2.

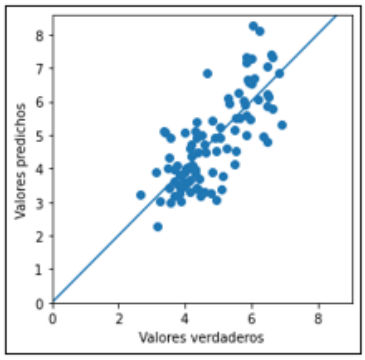

Gráfico D-44. Día 5 - retraso 3.

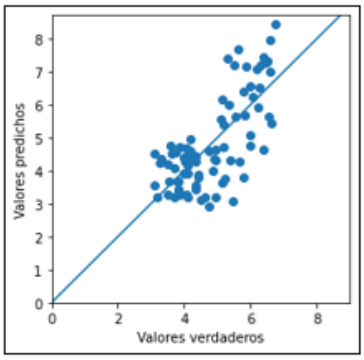

Gráfico D-45. Día 5 - retraso 4.

E. Anexo: Gráficas de dispersión de datos (NO_2) mediante Random Forest con diferentes retrasos por 5 días.

Gráfico D-6. Día 2 - retraso 0.

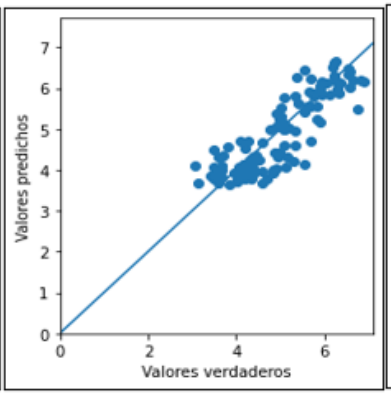

Gráfico D-7. Día 2 - retraso 1.

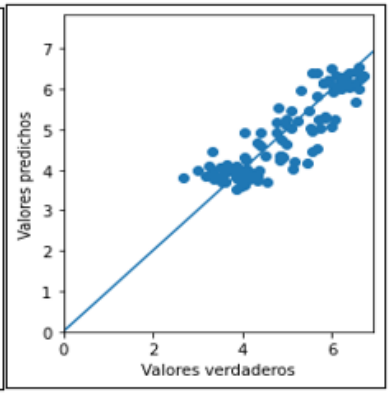

Gráfico D-8. Día 2 - retraso 2.

Gráfico D-9. Día 2 - retraso 3.

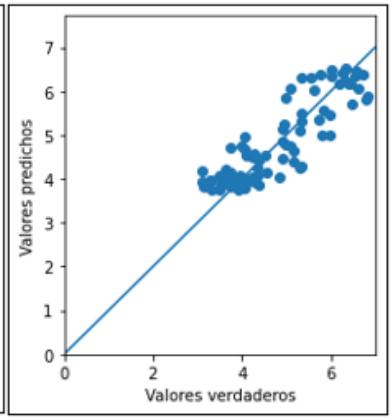

Gráfico D-10. Día 2 - retraso 4.

Gráfico D-16. Día 2 - retraso 0.

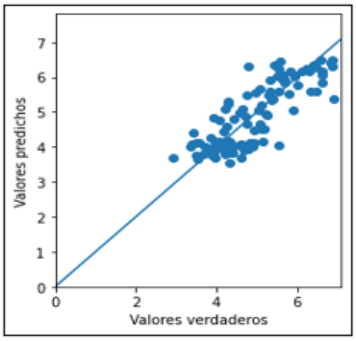

Gráfico D-17. Día 2 - retraso 1.

Gráfico D-18. Día 2 - retraso 2.

Gráfico D-19. Día 2 - retraso 3.

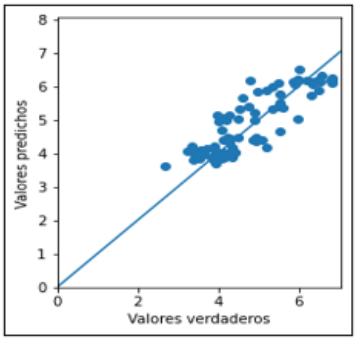

Gráfico D-20. Día 2 - retraso 4.

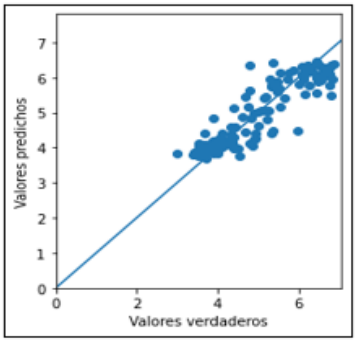

Gráfico D-26. Día 3 - retraso 0.

Gráfico D-27. Día 3 - retraso 1.

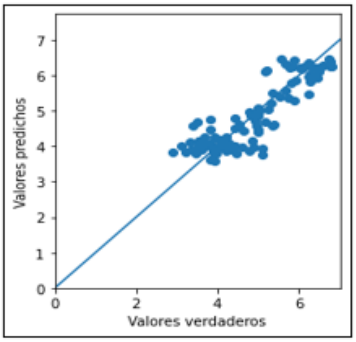

Gráfico D-28. Día 3 - retraso 2.

Gráfico D-29. Día 3 - retraso 3.

Gráfico D-30. Día 3 - retraso 4.

Gráfico D-36. Día 4 - retraso 0.

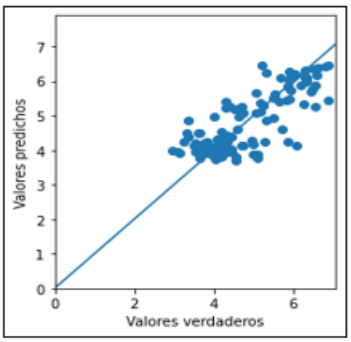

Gráfico D-37. Día 4 - retraso 1.

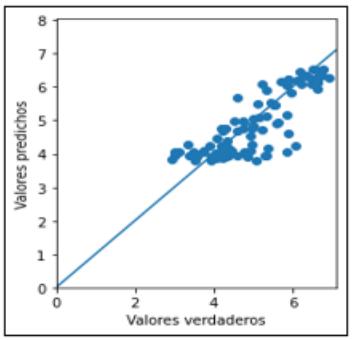

Gráfico D-38. Día 4 - retraso 2.

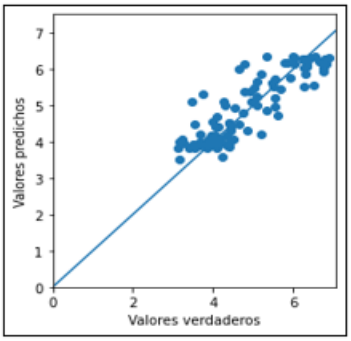

Gráfico D-39. Día 4 - retraso 3.

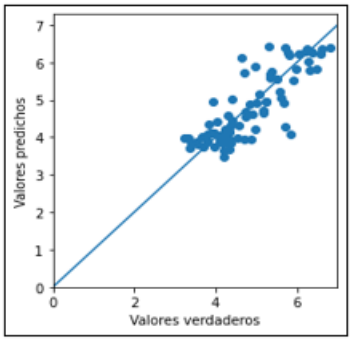

Gráfico D-40. Día 4 - retraso 4.

Gráfico D-46. Día 5 - retraso 0.

Gráfico D-47. Día 5 - retraso 1.

Gráfico D-48. Día 5 - retraso 2.

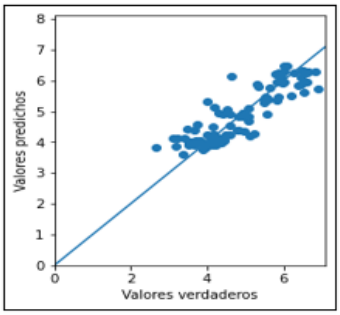

Gráfico D-49. Día 5 - retraso 3.

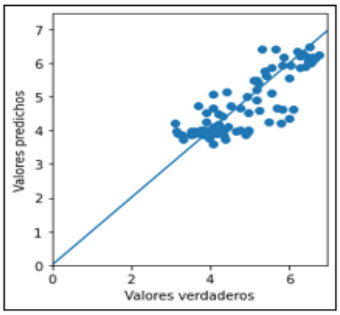

Gráfico D-50. Día 5 - retraso 4.

F. Anexo: Gráficas de dispersión de datos (SO_2) mediante Redes Neuronales Recurrentes con diferentes retrasos por 5 días.

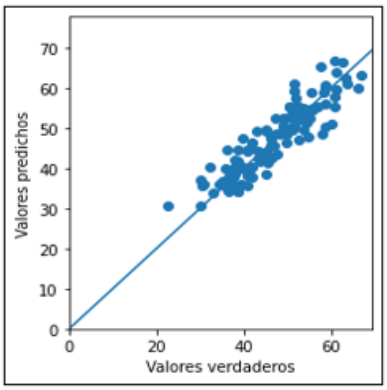

Gráfico F-1. Día 1 - retraso 0.

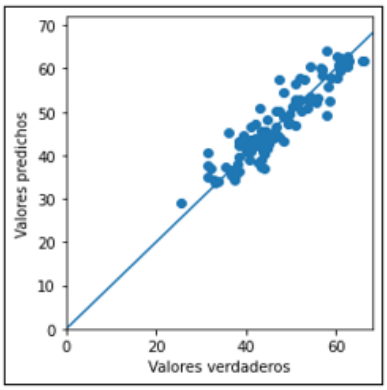

Gráfico F-2. Día 1 - retraso 1.

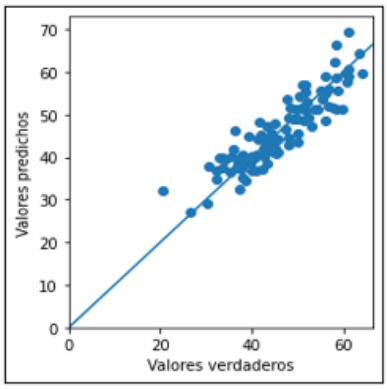

Gráfico F-3. Día 1 - retraso 2.

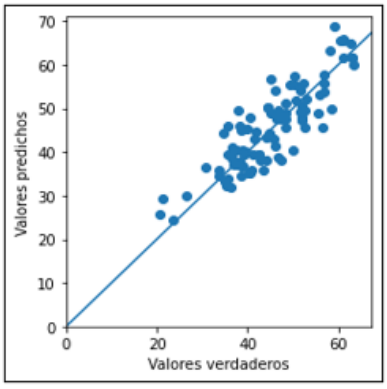

Gráfico F-4. Día 1 - retraso 3.

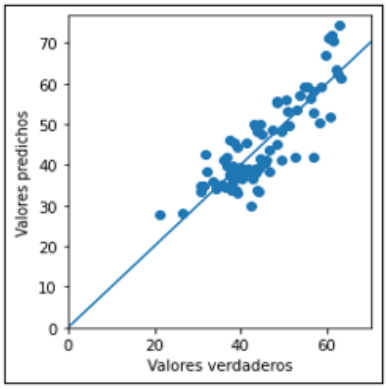

Gráfico F-5. Día 1 - retraso 4.

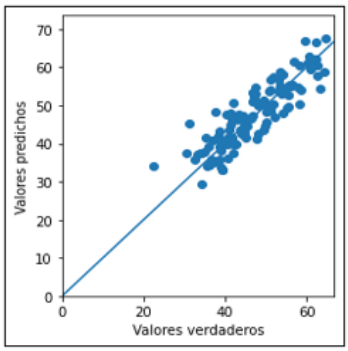

Gráfico F-6. Día 2 - retraso 0.

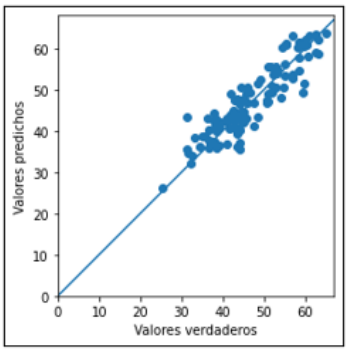

Gráfico F-7. Día 2 - retraso 1.

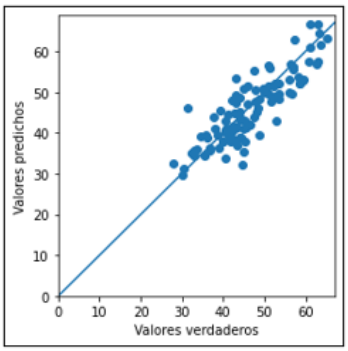

Gráfico F-8. Día 2 - retraso 2.

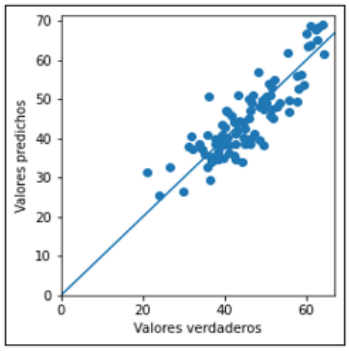

Gráfico F-9. Día 2 - retraso 3.

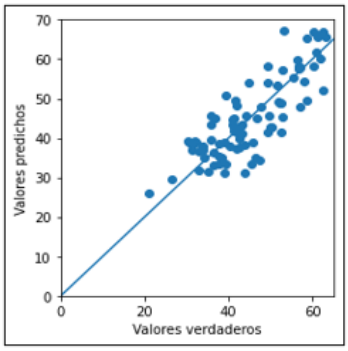

Gráfico F-10. Día 2 - retraso 4.

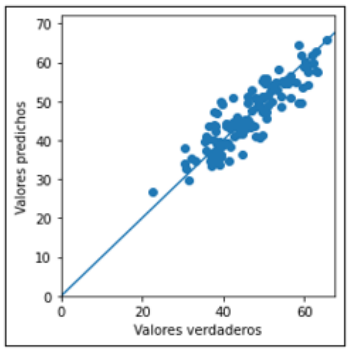

Gráfico F-6. Día 2 - retraso 0.

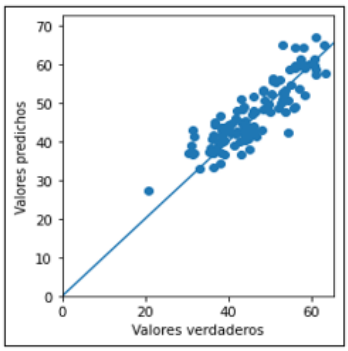

Gráfico F-7. Día 2 - retraso 1.

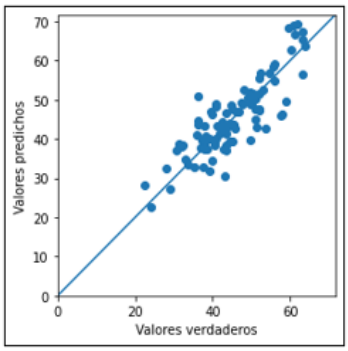

Gráfico F-8. Día 2 - retraso 2.

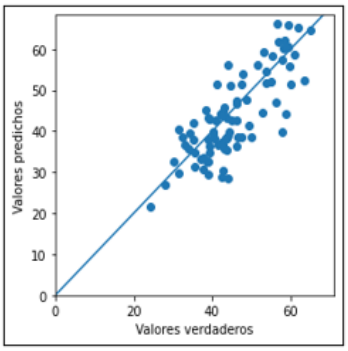

Gráfico F-9. Día 2 - retraso 3.

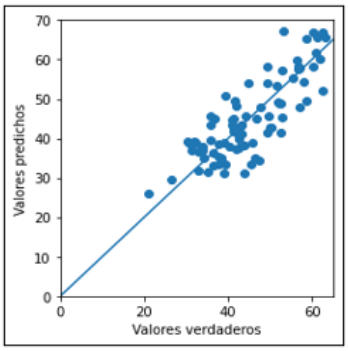

Gráfico F-10. Día 2 - retraso 4.

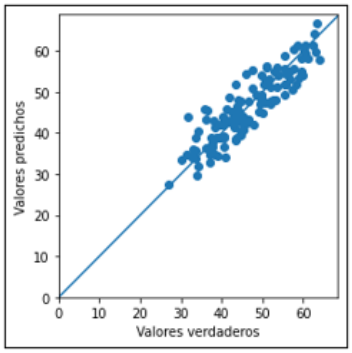

Gráfico F-16. Día 4 - retraso 0.

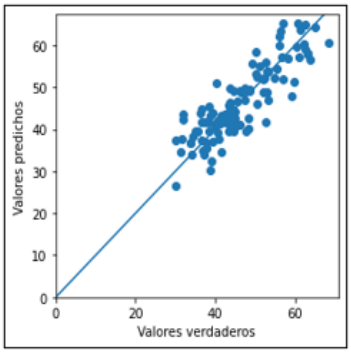

Gráfico F-17. Día 4 - retraso 1.

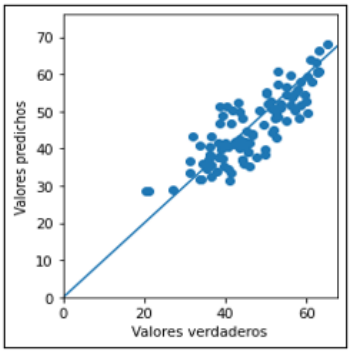

Gráfico F-18. Día 4 - retraso 2.

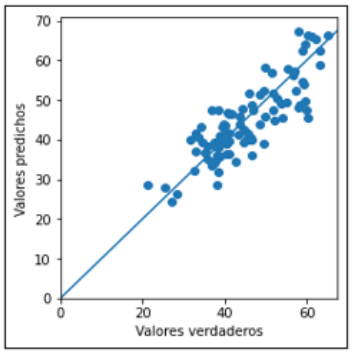

Gráfico F-19. Día 4 - retraso 3.

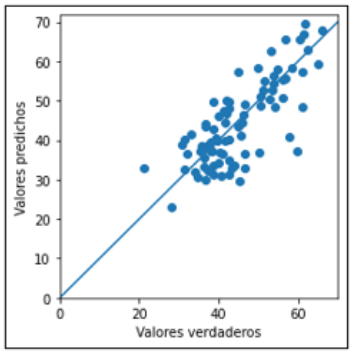

Gráfico F-20. Día 4 - retraso 4.

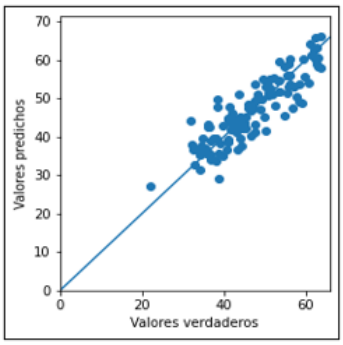

Gráfico F-21. Día 5 - retraso 0.

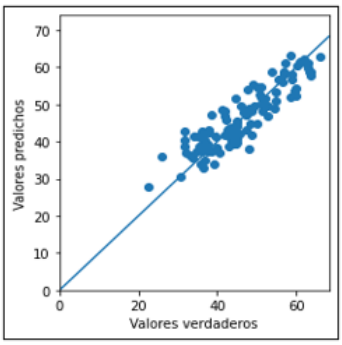

Gráfico F-22. Día 5 - retraso 1.

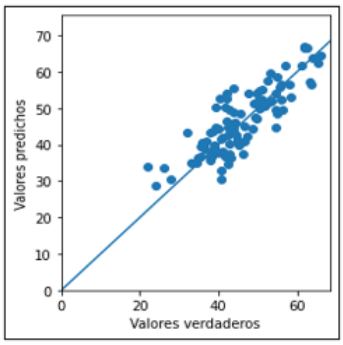

Gráfico F-23. Día 5 - retraso 2.

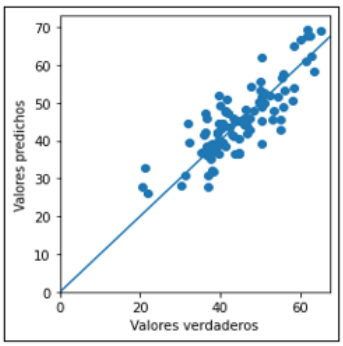

Gráfico F-24. Día 5 - retraso 3.

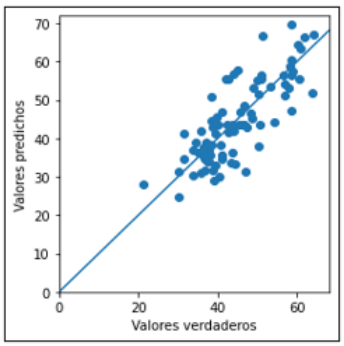

Gráfico F-25. Día 5 - retraso 4.

G. Anexo: Gráficas de dispersión de datos (SO_2) mediante Random Forest con diferentes retrasos por 5 días.

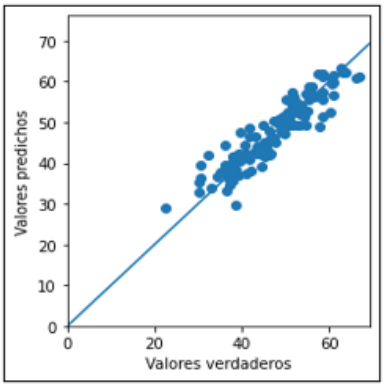

Gráfico G-1. Día 1 - retraso 0.

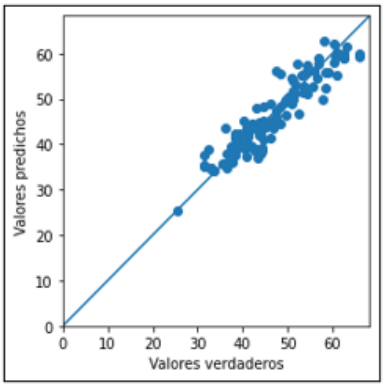

Gráfico G-2. Día 1 - retraso 1.

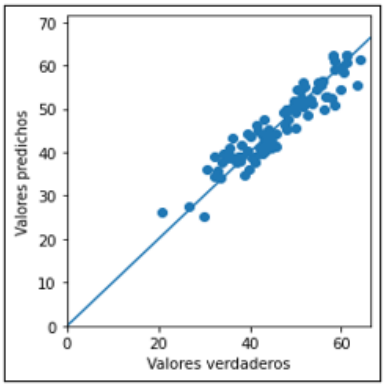

Gráfico G-3. Día 1 - retraso 2.

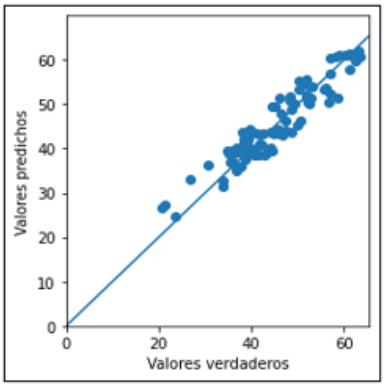

Gráfico G-4. Día 1 - retraso 3.

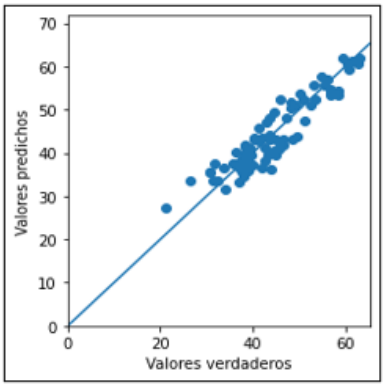

Gráfico G-5. Día 1 - retraso 4.

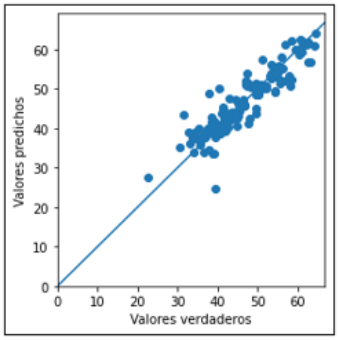

Gráfico G-6. Día 2 - retraso 0.

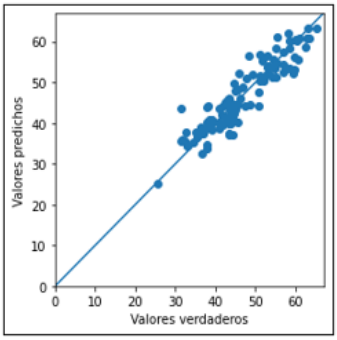

Gráfico G-7. Día 2 - retraso 1.

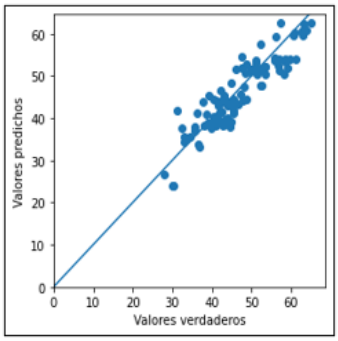

Gráfico G-8. Día 2 - retraso 2.

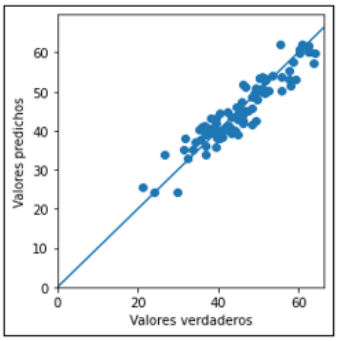

Gráfico G-9. Día 2 - retraso 3.

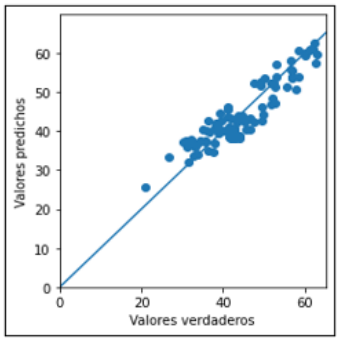

Gráfico G-10. Día 2 - retraso 4.

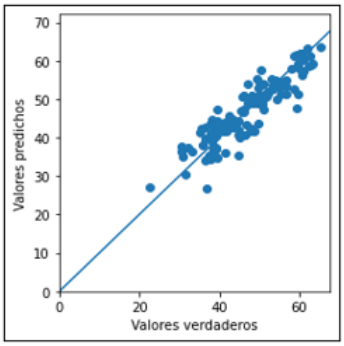

Gráfico G-11. Día 3 - retraso 0.

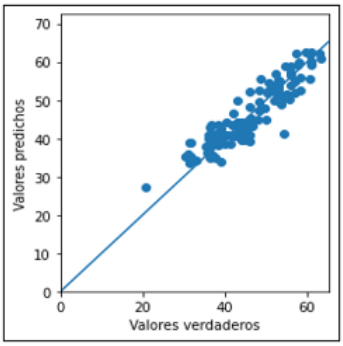

Gráfico G-12. Día 3 - retraso 1.

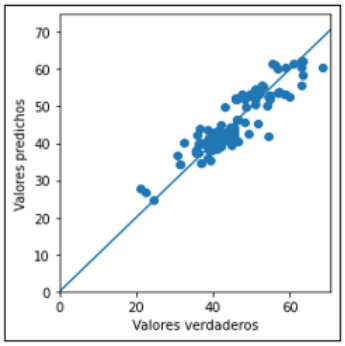

Gráfico G-13. Día 3 - retraso 2.

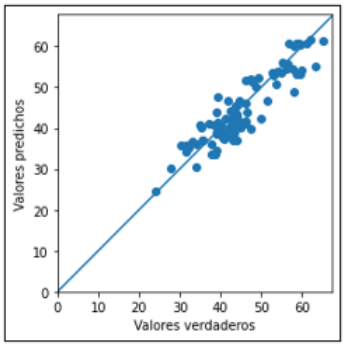

Gráfico G-14. Día 3 - retraso 3.

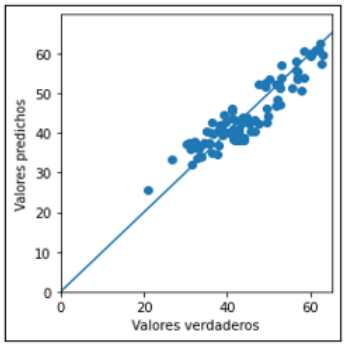

Gráfico G-15. Día 3 - retraso 4.

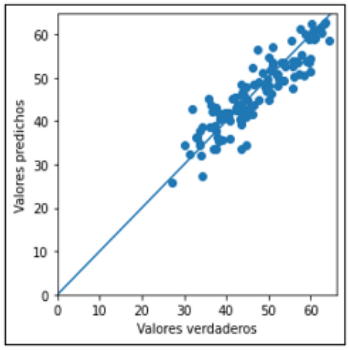

Gráfico G-16. Día 4 - retraso 0.

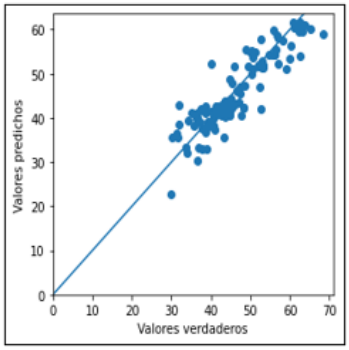

Gráfico G-17. Día 4 - retraso 1.

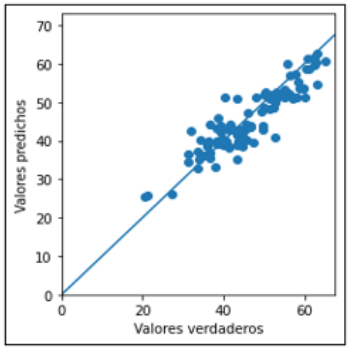

Gráfico G-18. Día 4 - retraso 2.

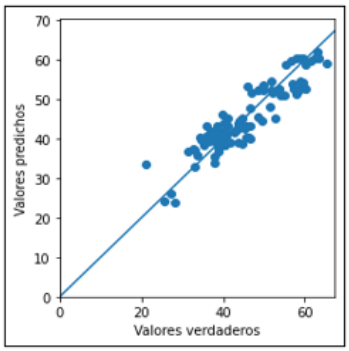

Gráfico G-19. Día 4 - retraso 3.

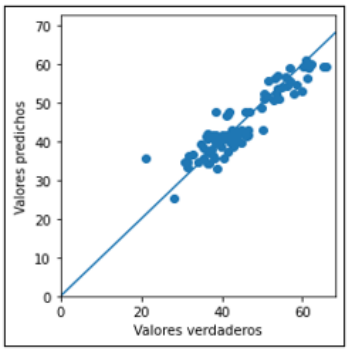

Gráfico G-20. Día 4 - retraso 4.

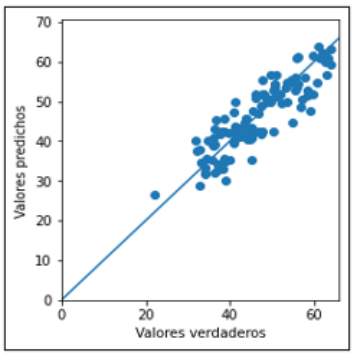

Gráfico G-21. Día 5 - retraso 0.

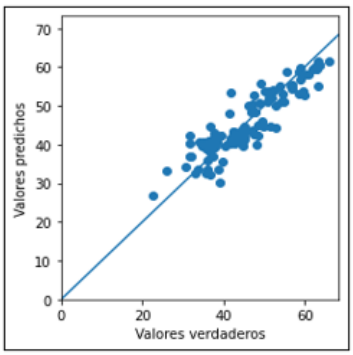

Gráfico G-22. Día 5 - retraso 1.

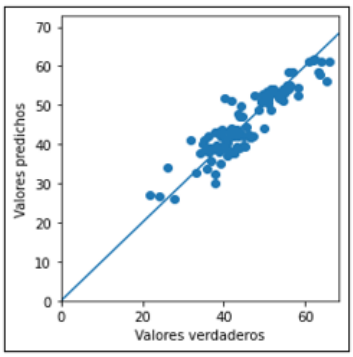

Gráfico G-23. Día 5 - retraso 2.

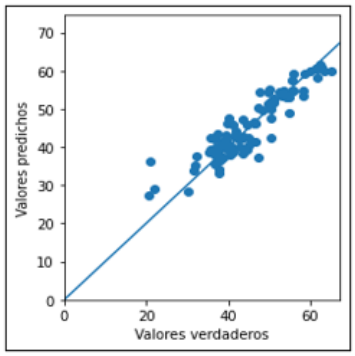

Gráfico G-24. Día 5 - retraso 3.

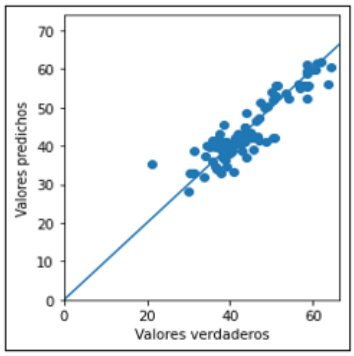

Gráfico G-25. Día 5 - retraso 4.

H. Anexo: Graficas de Relación de Pronostico de NO_2 mediante Redes Neuronales Recurrentes.

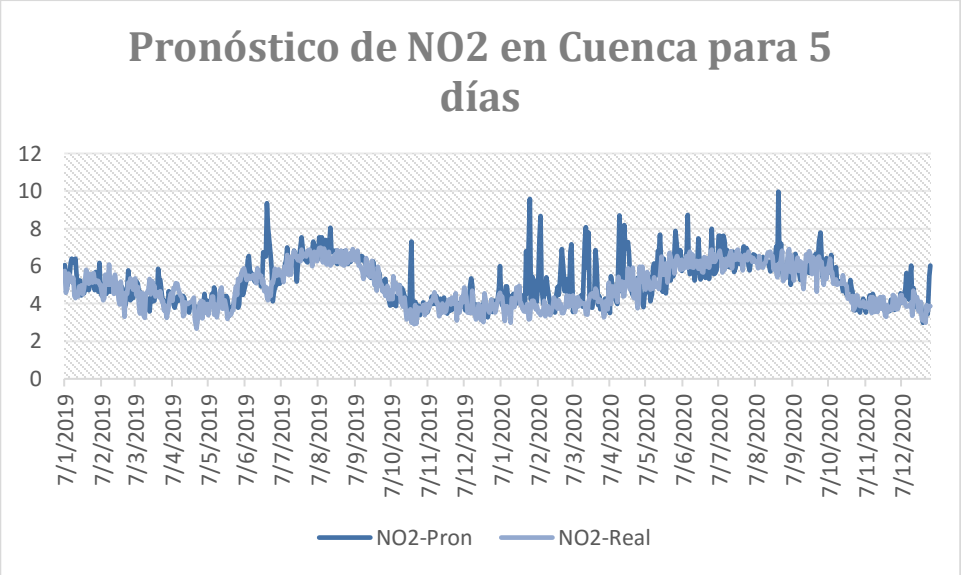

Fig. H-1. Pronóstico del NO_2 en Cuenca para 5 días.

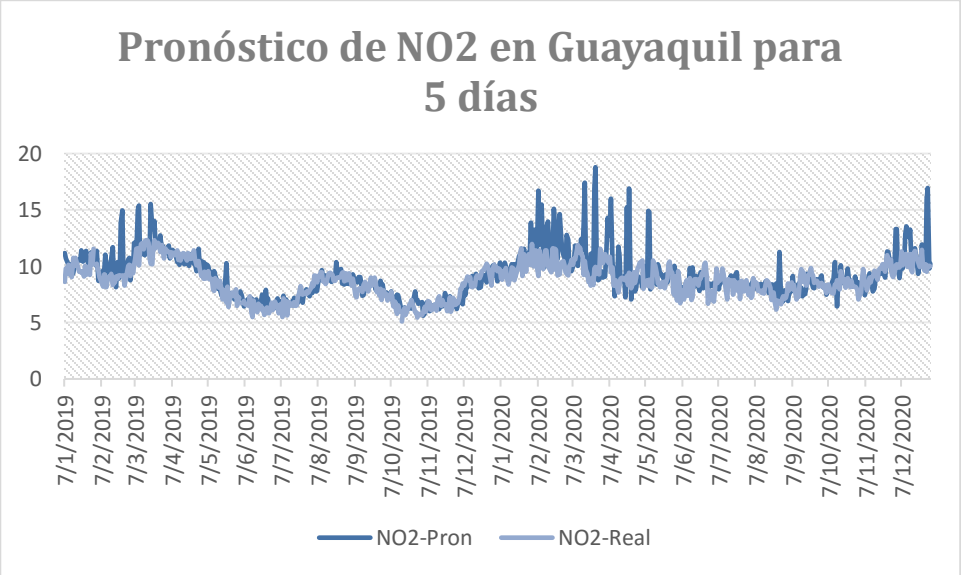

Fig. H-2. Pronóstico del NO_2 en Guayaquil para 5 días.

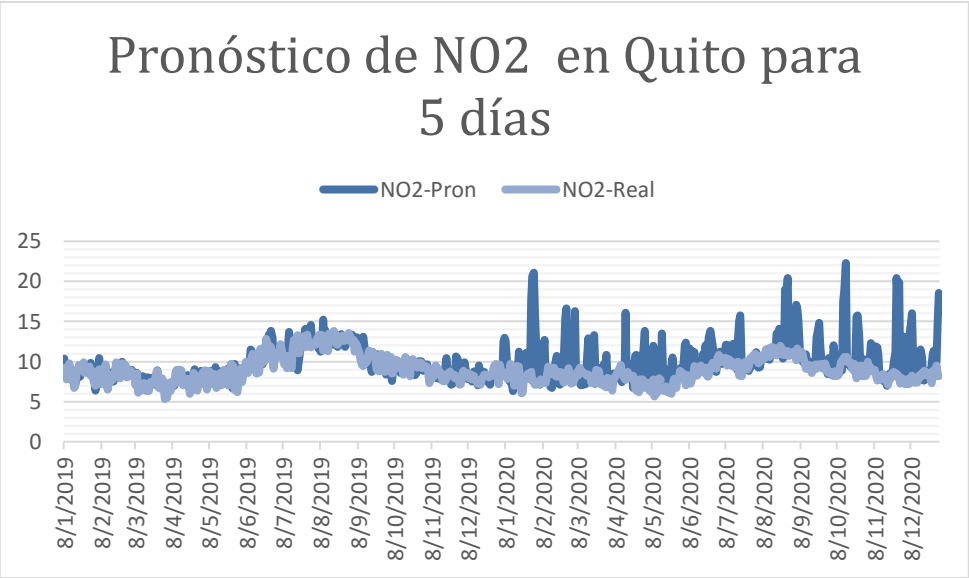

Fig. H-3. Pronóstico del NO_2 en Quito para 5 días.

I. Anexo: Gráficas de Relación de Pronóstico de SO_2 mediante Redes Neuronales Recurrentes

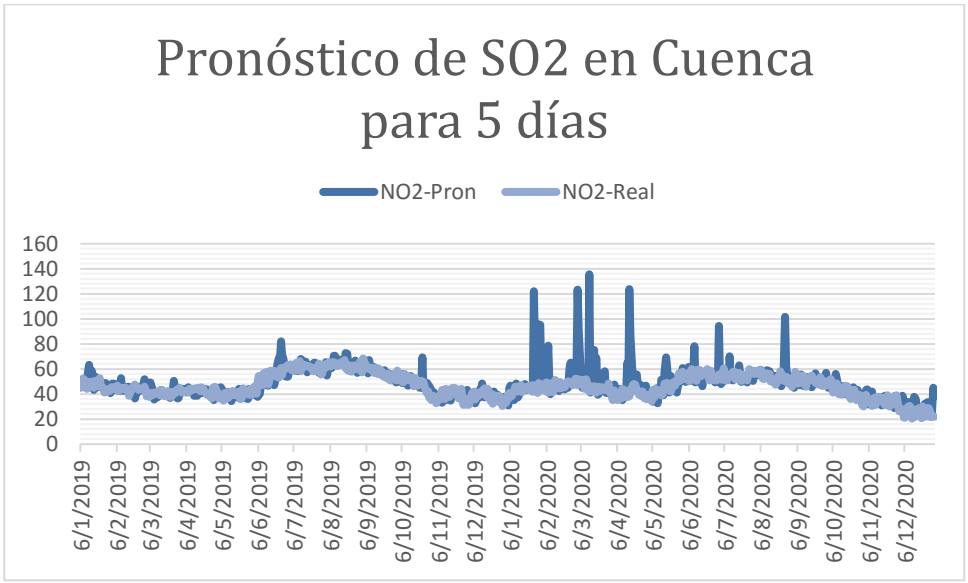

Fig. I-1. Pronóstico del NO_2 en Cuenca para 5 días.

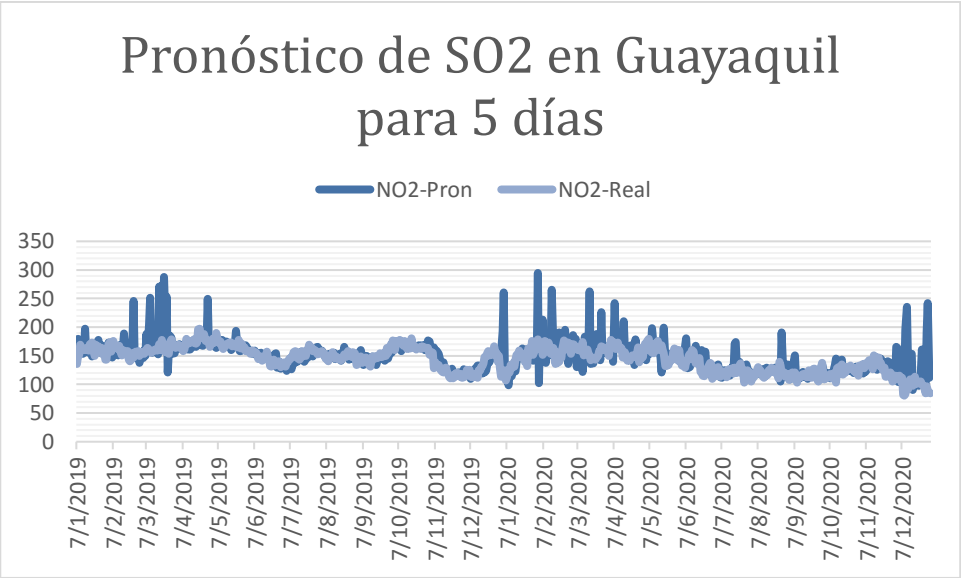

Fig. I-2. Pronóstico del SO_2 en Guayaquil para 5 días.

Fig. I-3. Pronóstico del SO_2 en Quito para 5 días.

J. Anexo: Gráficas de Relación de Pronóstico de NO_2 mediante Random Forest.

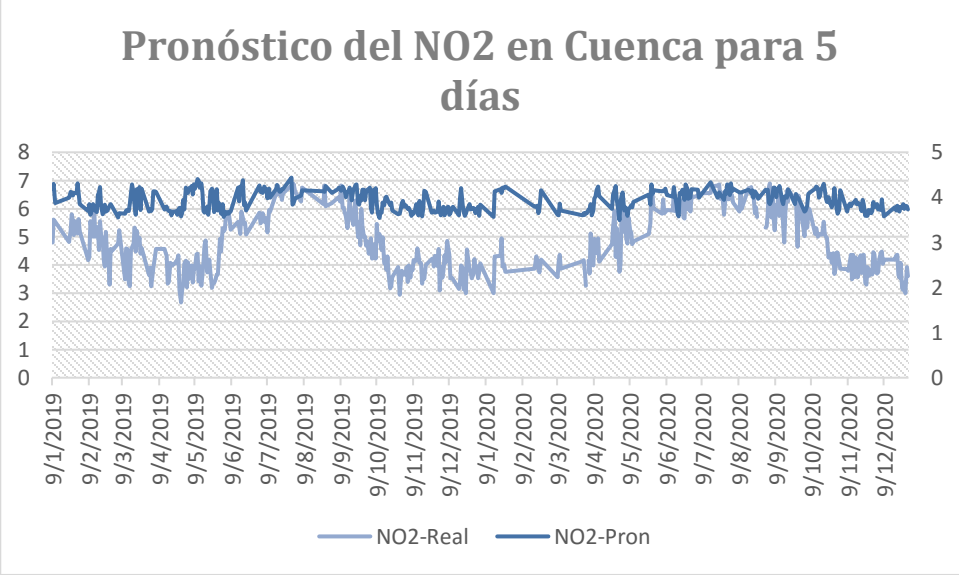

Fig. J-1. Pronóstico del NO_2 en Cuenca para 5 días.

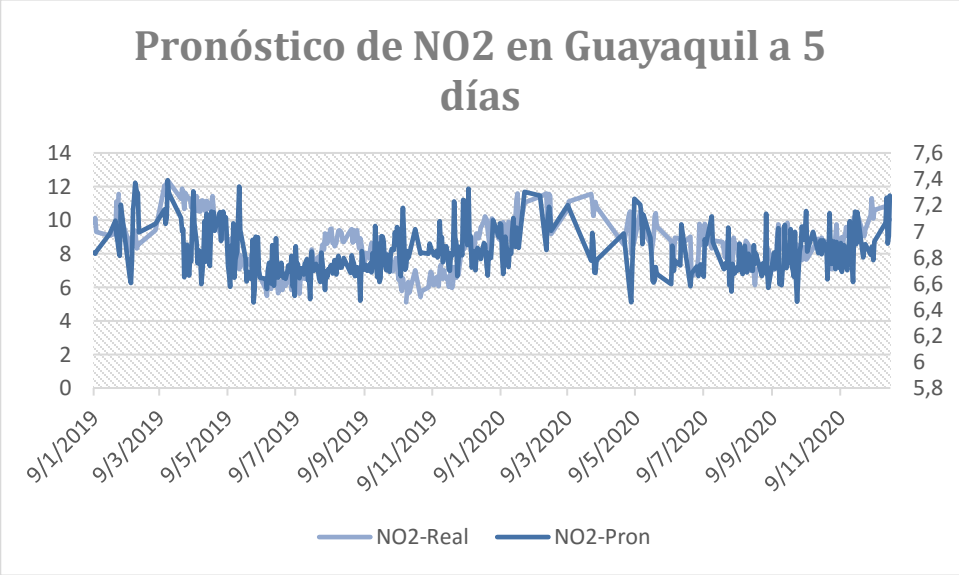

Fig. J-2. Pronóstico del NO_2 en Guayaquil para 5 días.

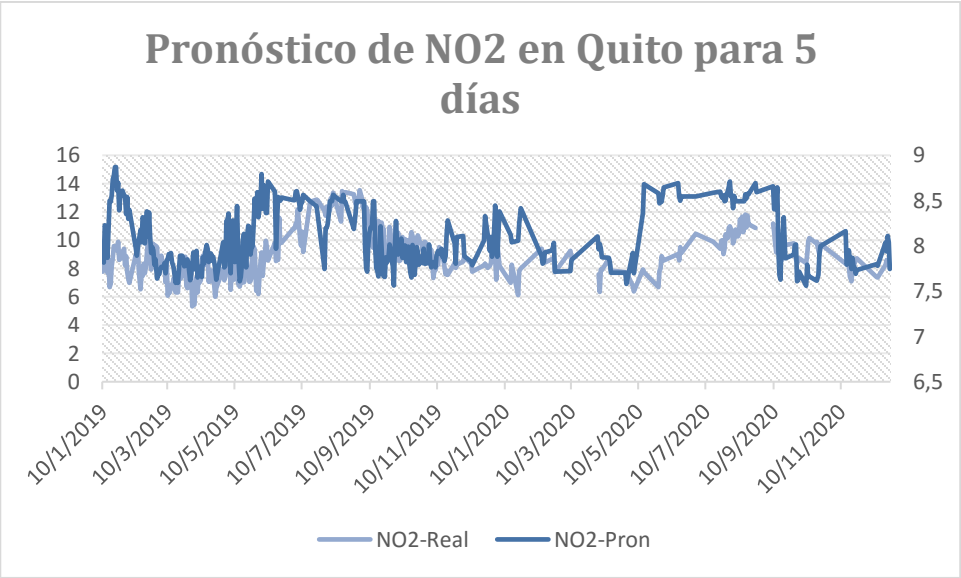

Fig. J-3. Pronóstico del NO_2 en Quito para 5 días.

K. Anexo: Gráficas de Relación de Pronóstico de SO_2 mediante Random Forest.

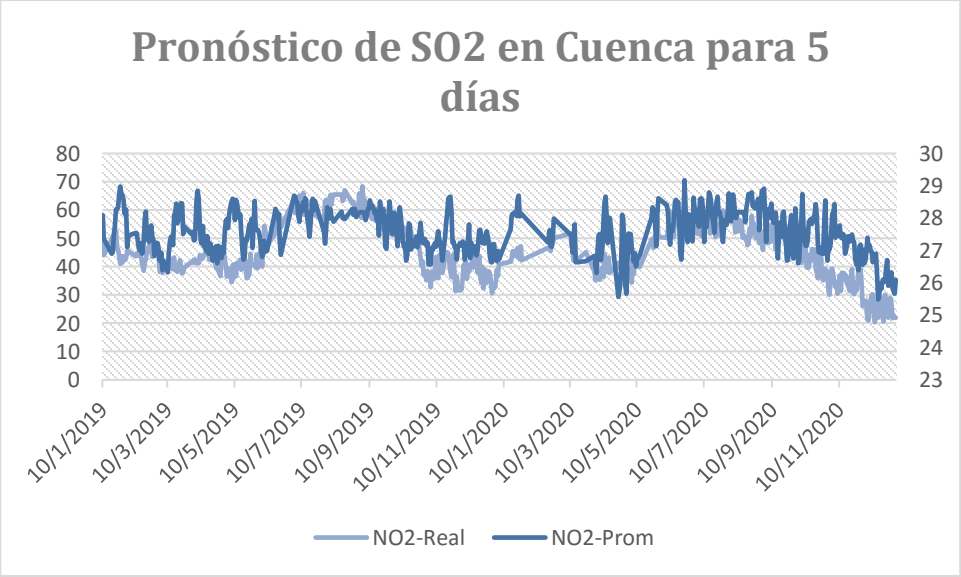

Fig. K-1. Pronóstico del NO_2 en Cuenca para 5 días.

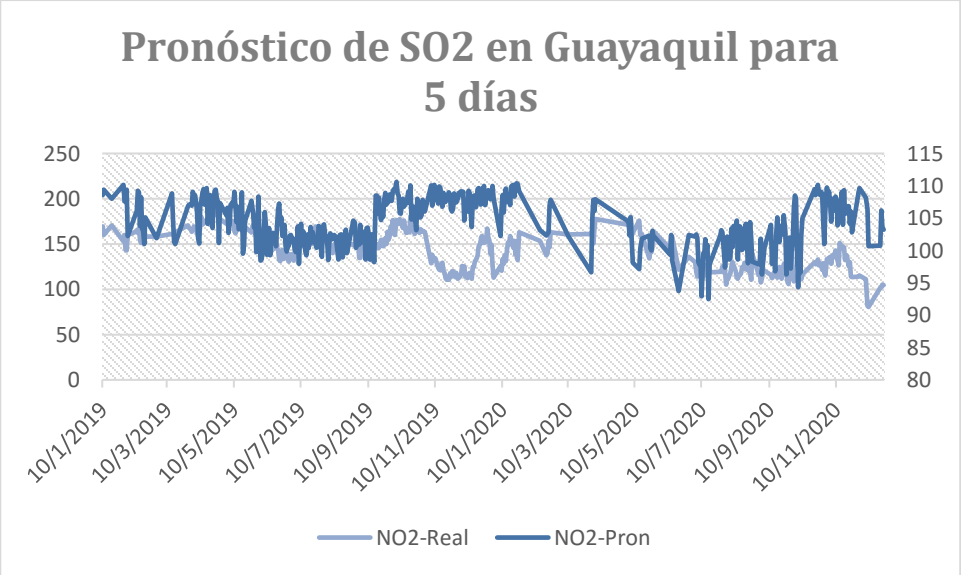

Fig. K-2. Pronóstico del SO_2 en Guayaquil para 5 días.

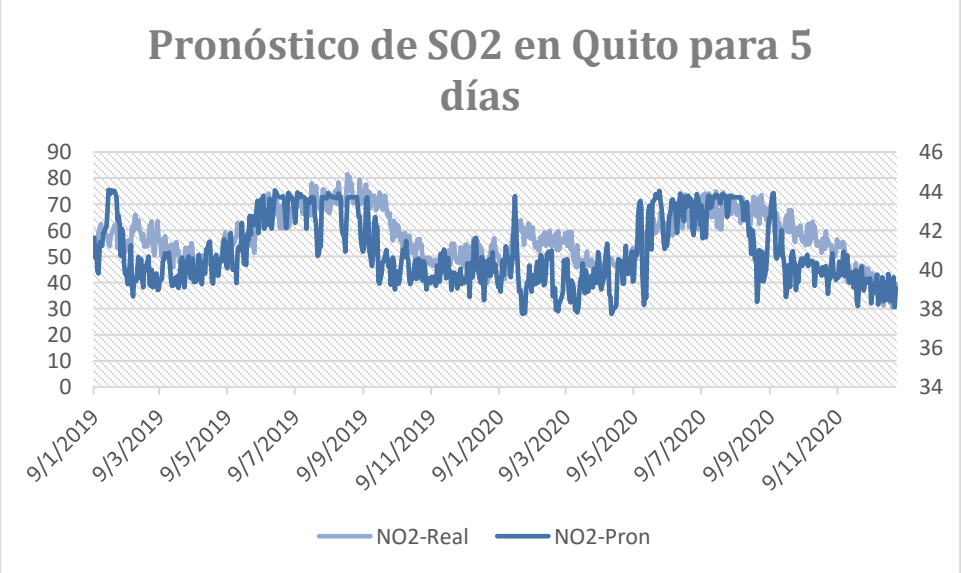

Fig. K-3. Pronóstico del SO_2 en Quito para 5 días.