

Desarrollo de un algoritmo para la predicción de tres parámetros ambientales

Development of an algorithm for the prediction of three environmental parameters

Johnatan Israel Corrales Bonilla¹ , Paola Gabriela Vinueza Naranjo¹ 

¹Universidad Nacional de Chimborazo, Riobamba – Ecuador

Correo de correspondencia: johnatan.corrales@unach.edu.ec, paolag.vinueza@unach.edu.ec

Información del artículo

Tipo de artículo:
Artículo original

Recibido:
02/09/2024

Aceptado:
06/10/2024

Publicado:
07/11/2024

Revista:
DATEH



Resumen

La anticipación concreta de parámetros medioambientales, generada con el uso de datos en tiempo real, es fundamental para optimizar el seguimiento y el control de la contaminación del aire y del cambio climático. En este artículo, se desarrolló un algoritmo basado en redes neuronales para predecir parámetros ambientales controlados y con datos adquiridos por sensores IoT. Se presenta una comparación y análisis de la precisión de cuatro modelos de Machine Learning: Regresión Lineal, Árbol de Decisión, Random Forest y Redes Neuronales. El objetivo principal de este manuscrito es predecir tres contaminantes críticos (ozono, monóxido de carbono y propano) y dos variables ambientales complementarias (temperatura y humedad). Los resultados revelan que los modelos de predicción de Árbol de Decisión y Random Forest son especialmente precisos para la variable humedad, mientras que las Redes Neuronales destacan en la predicción de los niveles de ozono. Este estudio demuestra la capacidad de los modelos avanzados de aprendizaje automático para manejar datos ambientales complejos. Estos hallazgos importantes se proponen como un punto de partida para ser evaluados de forma detallada en diferentes escenarios con implicaciones para la gestión ambiental y la salud pública.

Palabras clave: *Predicción Ambiental, Redes Neuronales, Sensores IoT, Modelos de Predicción, Aprendizaje Automático.*

Abstract

The concrete anticipation of environmental parameters, generated with the use of real-time data, is essential to optimize the monitoring and control of air pollution and climate change. In this paper, an algorithm based on neural networks was developed to predict controlled environmental parameters and with data acquired by IoT sensors. A comparison and analysis of the accuracy of four Machine Learning models Linear Regression, Decision Tree, Random Forest and Neural Networks is presented. The main objective of this manuscript is to predict three critical pollutants (ozone, carbon monoxide and propane) and two complementary environmental variables (temperature and humidity). The results reveal that Decision Tree and Random Forest prediction models are especially accurate for the humidity variable, while Neural Networks excel in the prediction of ozone levels. This study demonstrates the ability of advanced machine learning models to handle complex environmental data. These important findings are proposed as a starting point to be evaluated in detail in different scenarios with implications for environmental management and public health.

Keywords: *Environmental Prediction, Neural Networks, IoT Sensors, Prediction Models, Machine Learning*

Forma sugerida de citar (APA): López-Rodríguez, C. E., Sotelo-Muñoz, J. K., Muñoz-Venegas, I. J. y López-Aguas, N. F. (2024). Análisis de la multidimensionalidad del brand equity para el sector bancario: un estudio en la generación Z. Retos Revista de Ciencias de la Administración y Economía, 14(27), 9-20. <https://doi.org/10.17163/ret.n27.2024.01>.

INTRODUCCIÓN

Las Redes Neuronales Artificiales (ANN) han revolucionado el campo de la predicción ambiental al ofrecer herramientas robustas para modelar y pronosticar diversos parámetros. Las ANN han demostrado ser efectivas en el aprendizaje de patrones complejos a partir de grandes volúmenes de datos, haciendo posible la predicción precisa de contaminantes atmosféricos como el

ozono, el monóxido de carbono y compuestos orgánicos volátiles (VOCs) (Alwan et al., 2023; Pandya et al., 2020; Shukla et al., 2023). Además, las ANN se aplican exitosamente en la predicción de parámetros de calidad del agua, tales como turbidez, salinidad y concentración de nutrientes, siendo cruciales para la gestión y monitoreo de recursos hídricos (Alotaibi et al., 2022; Baranitharan et al., 2021). En el ámbito climático, las ANN, han mostrado

eficacia en la predicción de variables como la temperatura, precipitación y radiación solar, facilitando el manejo de grandes conjuntos de datos y múltiples variables de entrada (Rao et al., 2022; Shukla et al., 2023).

Dentro de las ANN existen subcategorías especializadas que amplían sus aplicaciones. Las Redes Neuronales Feedforward (RNF), también conocidas como Feedforward Neural Networks (FFNN), son apreciadas por su simplicidad y eficiencia computacional en tareas que no requieren dependencias temporales explícitas (Frolov et al., 2023; Shu et al., 2024). Aunque, las RNF se destacan en la captura de no linealidades complejas, su aplicabilidad es más limitada en comparación con las Redes Neuronales Recurrentes (RNN) y las Redes de Memoria a Largo Plazo (LSTM), debido a su incapacidad para modelar dependencias temporales complejas (Cen et al., 2023; Popli et al., 2024). Las RNN y LSTM son ideales para datos con dependencias temporales, como el pronóstico de la calidad del aire y el caudal de ríos, aunque presentan una mayor complejidad y tiempo de entrenamiento (Ding et al., 2018; Suleman & Shridevi, 2022). Las Redes Neuronales Convolucionales (CNN), diseñadas principalmente para datos espaciales, ofrecen ventajas en el análisis de imágenes satelitales y el monitoreo de la salud de los ecosistemas, aunque pueden ser excesivas para datos no visuales (J. Chen et al., 2023; H. Yang et al., 2023).

La comprensión precisa de los parámetros ambientales es crucial para enfrentar desafíos contemporáneos como el cambio climático, la contaminación del aire y la gestión de recursos hídricos (Wu et al., 2020). Los datos generados por sensores IoT (Internet de las Cosas) proporcionan una oportunidad sin precedentes para monitorear y analizar estos parámetros en tiempo real y a gran escala (Manocha et al., 2024). Sin embargo, la extracción de información significativa de estos datos complejos requiere técnicas avanzadas de análisis (Sathyaseelan & Sarathambekai, 2021). En este contexto, surge la necesidad de desarrollar algoritmos eficientes y precisos para la predicción de parámetros ambientales utilizando datos de sensores IoT, con aplicaciones potenciales en la agricultura, la industria, la gestión urbana y la salud pública (Huu, 2024). Por ejemplo, una predicción precisa de la calidad del aire puede asistir a las autoridades en la implementación de medidas para reducir la contaminación y proteger la salud pública (Y. Chen et al., 2020; J. Yang et al., 2022).

En este manuscrito se propone un enfoque integral para la predicción de parámetros ambientales, como: ozono, monóxido de carbono, propano, temperatura y humedad, utilizando datos generados por sensores IoT. La red neuronal utilizada en este manuscrito es ANN, de tipo Perceptrón Multicapa (MLP), que forma parte de las

FFNN, presenta una estructura de múltiples capas, con una capa de entrada equivalente al número de características relevantes, tres capas ocultas con 128, 64 y 32 nodos respectivamente, y una capa de salida con un solo nodo para la predicción continua. Las capas ocultas emplean la función de activación ReLU, regularización L2 y Dropout para prevenir el sobreajuste. La red se compila utilizando el optimizador Adam y la función de pérdida `mean_squared_error`, adecuada para problemas de regresión. Esta configuración permite a la red modelar relaciones complejas entre las variables y predecir valores continuos a partir de los datos de entrada.

Además, se evalúa cuatro modelos de aprendizaje automático: Regresión Lineal (LR), Árbol de Decisión (DT), Random Forest (RF) y Redes Neuronales (NN), con el objetivo de comparar su precisión, eficiencia y adaptabilidad a diversas condiciones ambientales. En este manuscrito no solo se examina el rendimiento de los modelos en términos de precisión y eficiencia, sino que también aborda aspectos clave como la integración de datos de sensores IoT, la calidad de los datos, la selección de características y el diseño del modelo.

MATERIALES Y MÉTODOS

Diseño Experimental

En este manuscrito se desarrolla un algoritmo basado en redes neuronales para la predicción de parámetros ambientales utilizando datos recopilados por sensores. Se emplearon técnicas de modelado predictivo para estimar variables ambientales clave como Temperatura, Humedad, concentraciones de gases como, Monóxido de Carbono, Ozono y Propano.

Materiales

Dataset: Se utilizó el dataset de los registros de mediciones ambientales obtenidos por los sensores de Ozono, Propano, Monóxido de Carbono, Temperatura y Humedad. Los datos se recopilaron a lo largo de 6 días con un intervalo de muestreo de 5 segundos y se organizaron en un formato adecuado para el análisis y entrenamiento de modelos, consiguiendo con esto 212126 datos normalizados para el estudio.

Sensores: Se utilizaron los siguientes sensores para la medición de variables ambientales:

- **MQ-131 (Ozono):** Este sensor fue utilizado para medir la concentración de ozono en el ambiente. El MQ-131 tiene la capacidad de detectar niveles bajos y altos de ozono, proporcionando datos cruciales para la evaluación de la calidad del aire.
- **MQ-7 (Monóxido de Carbono):** El sensor MQ-7 se empleó para la detección de monóxido de carbono (CO). Este sensor es especialmente sensible a bajas concentraciones de CO, lo que es vital para

monitorear la contaminación por monóxido de carbono y evaluar riesgos potenciales para la salud.

- **MQ-2 (Propano):** Este sensor fue utilizado para la medición de gases como el propano, metano y otros gases combustibles. El MQ-2 es capaz de detectar la presencia de estos gases en el aire, lo que es importante para la prevención de fugas y la seguridad en entornos industriales y residenciales.
- **Sensor de Temperatura:** Este sensor se utilizó para registrar la temperatura ambiental. La medición precisa de la temperatura es fundamental para entender las condiciones climáticas y su impacto en las variables ambientales que se están estudiando.
- **Sensor de Humedad:** El sensor de humedad proporcionó datos sobre la cantidad de vapor de agua en el aire. Esta medición es esencial para analizar cómo la humedad puede influir en otros parámetros ambientales, como la calidad del aire y la concentración de gases.

Hardware computacional: La tabla 1, detalla las especificaciones del hardware del sistema utilizado, que incluye el procesador, la memoria, y la tarjeta de video. Estos componentes son fundamentales para el rendimiento general del sistema, especialmente en tareas que requieren un procesamiento intensivo como el aprendizaje automático y el análisis de datos.

Componente	Especificaciones
Procesador	Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz
Memoria	8,00 GB (7,89 GB utilizable)
Tarjeta de Video	Detalles a continuación
Número de Núcleos de Procesamiento	8704 núcleos CUDA
Núcleos Tensor	Sí, con soporte para operaciones de matrices optimizadas para aprendizaje profundo
Memoria de Video	10 GB GDDR6X
Ancho de Banda de Memoria	760 GB/s
Compatibilidad con Tecnologías de Aceleración	Soporte completo para CUDA, cuDNN, TensorFlow y PyTorch

Tabla 1. Características de hardware del sistema.

Entorno de desarrollo: Se ha empleado Python 3 como lenguaje de programación principal, complementándolo con Jupyter Notebook como entorno de desarrollo para facilitar un enfoque interactivo y eficiente en el análisis de datos y el modelado. En la tabla 2, se detallan las bibliotecas de Python utilizadas, cada una desempeñando un papel crucial en el flujo de trabajo. Estas bibliotecas incluyen pandas y numpy para la manipulación y análisis de datos, matplotlib y seaborn para la visualización de datos, joblib para la serialización de objetos, y sklearn para

el aprendizaje automático. También, se utilizó, tensorflow para la construcción y entrenamiento de modelos de redes neuronales profundas.

Biblioteca	Descripción
pandas	Biblioteca para la manipulación y análisis de datos. Permite la carga de datos, manipulación de DataFrames, etc.
numpy	Biblioteca para operaciones matemáticas y manipulación de arreglos. Proporciona soporte para arrays multidimensionales y funciones matemáticas.
matplotlib	Biblioteca para la creación de gráficos y visualizaciones. Permite generar gráficos de dispersión, líneas, barras, entre otros.
seaborn	Biblioteca para la visualización de datos basada en Matplotlib. Facilita la creación de gráficos estadísticos con estilo y funcionalidad adicionales.
Joblib	Biblioteca para la serialización de objetos en Python. Utilizada aquí para guardar modelos y escaladores en archivos.
sklearn	Biblioteca para aprendizaje automático. Incluye herramientas para preprocesamiento de datos, selección de modelos, evaluación y más.
tensorflow	Biblioteca de código abierto para la construcción y entrenamiento de modelos de aprendizaje profundo. Se utiliza aquí para crear y entrenar una red neuronal.

Tabla 2. Bibliotecas de Python usadas.

Modelos predictivos

Se utilizaron cuatro modelos de aprendizaje automático para predecir las variables seleccionadas en este estudio. A continuación, se describen las características principales de cada uno.

El primer modelo empleado fue (LR), un enfoque clásico que asume una relación lineal entre las características independientes y la variable objetivo.

El segundo modelo fue el (DT), una técnica que divide iterativamente los datos en subconjuntos según reglas de decisión, basadas en qué características optimizan la predicción de la variable objetivo. Este modelo es capaz de capturar relaciones no lineales complejas, lo que lo convierte en una opción más flexible que la regresión lineal.

El tercer modelo fue el (RF), un método de ensamble que combina múltiples árboles de decisión. Cada uno de estos árboles se entrena en diferentes subconjuntos de datos y características, lo que reduce la varianza y mejora la precisión del modelo.

Finalmente, se empleó una Red Neuronal (Perceptrón Multicapa, NN), compuesta por tres capas ocultas con 128,

64 y 32 neuronas, respectivamente. Cada capa utilizó la función de activación ReLU, que es eficaz para capturar relaciones no lineales complejas. Para evitar el sobreajuste, se implementaron técnicas de regularización L2 y Dropout en cada capa oculta, donde el 30% de las neuronas se desactivaron aleatoriamente durante el entrenamiento. El modelo fue entrenado, utilizando el optimizador Adam con una tasa de aprendizaje de 0.001, durante 200 épocas.

Para todos estos modelos, se entrenó utilizando las tres características más importantes, seleccionadas previamente mediante el análisis de importancia de características del modelo de Random Forest. La evaluación del modelo se realizó calculando el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2) en el conjunto de prueba, permitiendo comparar los resultados de la red neuronal con los otros modelos.

Procedimiento

En la figura 1, se identifica las etapas fundamentales en el desarrollo de un modelo de predicción ambiental utilizando ANN, específicamente de tipo Perceptrón Multicapa (MLP). Primero, en la recolección de datos, se recopilan datos generados por sensores IoT que miden parámetros ambientales como ozono, monóxido de carbono, propano, temperatura y humedad. Estos sensores proporcionan datos en tiempo real y a lo largo del tiempo, lo que permite capturar patrones y tendencias en las variables ambientales, siendo esenciales para el éxito del modelo de predicción, ya que aseguran que los datos sean representativos y abarquen diversas condiciones ambientales.

Luego, en el preprocesamiento de datos, los datos obtenidos se limpian y transforman para asegurar su calidad y adecuación al modelo. Este paso incluye la eliminación de valores nulos o atípicos, la normalización o estandarización de los datos, y la transformación de variables categóricas en numéricas si es necesario, eliminando ruidos e inconsistencias que puedan afectar el rendimiento del modelo y asegurando que los datos sean aptos para el análisis y el modelado. Posteriormente, en la selección de características, se identifican y seleccionan las variables más relevantes que influirán en la precisión del modelo. La selección de características es un proceso crítico para reducir la dimensionalidad del conjunto de datos y eliminar información redundante o irrelevante.

Técnicas como la matriz de correlación, la importancia de características usando modelos de árbol (e.g., RF), y métodos de selección automática se utilizan para este propósito, mejorando la eficiencia del modelo y

concentrándose en las variables que tienen mayor impacto en la predicción.

En el desarrollo del modelo de una red neuronal, se diseña y configura una ANN de tipo Perceptrón Multicapa (MLP). La estructura de la red incluye una capa de entrada con el número de características relevantes seleccionadas, tres capas ocultas con 128, 64 y 32 nodos respectivamente, y una capa de salida con un solo nodo para la predicción continua. Las capas ocultas emplean la función de activación ReLU (Rectified Linear Unit), regularización L2 para evitar el sobreajuste, y Dropout para mejorar la generalización del modelo. La red se compila utilizando el optimizador Adam y la función de pérdida `mean_squared_error`, adecuada para problemas de regresión. Esta configuración permite que la red neuronal capture complejas relaciones no lineales entre las variables de entrada y la variable objetivo.

En la etapa de entrenamiento y evaluación del modelo, el modelo se entrena utilizando los datos disponibles, ajustando sus pesos para minimizar la función de pérdida y aprendiendo a partir de los datos de entrenamiento. Después del entrenamiento, se evalúa el rendimiento del modelo en un conjunto de datos de prueba que no se utilizó durante el entrenamiento, utilizando métricas de rendimiento como el error cuadrático medio (MSE) y el coeficiente de determinación R^2 para evaluar la precisión del modelo. Esta fase permite ajustar los parámetros del modelo y validar su capacidad para generalizar a datos no vistos previamente.

Finalmente, en la validación y comparación con otros modelos, los resultados obtenidos se validan y se comparan con otros modelos de aprendizaje automático, como LR, DT y RF. Este análisis comparativo permite evaluar la precisión, eficiencia y adaptabilidad de cada modelo bajo diversas condiciones ambientales, ayudando a identificar el modelo más adecuado para la tarea de predicción específica, considerando tanto el rendimiento como la capacidad del modelo para manejar diferentes escenarios. La validación y comparación aseguran que los resultados obtenidos sean confiables y que el modelo seleccionado sea el más efectivo para las aplicaciones previstas.

- 1. Recolección de Datos:**
 - Sensores utilizados: MQ-131, MQ-7, MQ-2, Temperatura, Humedad
 - Intervalo de muestreo: 5 segundos
 - Duración del período de recolección: 6 días
- 2. Preprocesamiento de Datos:**
 - Ordenamiento y pivoteo de datos
 - Limpieza de datos crudos para eliminar valores atípicos y datos faltantes
 - Normalización de datos para escalar a un rango uniforme
- 3. Selección de Características:**
 - Análisis de correlación
 - Uso de Random Forest para seleccionar características relevantes
- 4. Desarrollo del Modelo de Red Neuronal:**
 - Utilización de TensorFlow y Keras
 - Arquitectura del modelo:
 - División de datos: 80% entrenamiento, 20% prueba
 - Capas densas con activación ReLU
 - Capa de salida para predicción numérica
- 5. Entrenamiento y Evaluación del Modelo:**
 - Entrenamiento con datos históricos
 - Evaluación con métricas:
 - Error Cuadrático Medio (MSE)
 - Coeficiente de Determinación (R^2)
- 6. Validación y Comparación con Otros Modelos:**
 - Comparación con modelos de:
 - Regresión Lineal
 - Árboles de Decisión
 - Bosques Aleatorios
 - Evaluación de rendimiento con las mismas métricas

Figura 1. Metodología de Investigación

Diseño y Entrenamiento de la Red Neuronal Artificial (MLP)

La figura 2 muestra el diseño de una red neuronal tipo Perceptrón Multicapa (MLP) para predecir parámetros ambientales como el ozono, monóxido de carbono y temperatura.

La red inicia con una capa de entrada que recibe estas características relevantes. Luego, cuenta con tres capas ocultas que utilizan la activación ReLU: la primera con 128 neuronas, la segunda con 64 y la tercera con 32 neuronas. Además, se aplican técnicas para mejorar el rendimiento y evitar el sobreajuste, como la regularización L2 y una tasa de Dropout del 30% en cada capa.

En la capa de salida, un único nodo genera las predicciones continuas de los niveles de ozono. El modelo se entrena utilizando el optimizador Adam y la función de pérdida `mean_squared_error` (MSE), durante 200 épocas. Para evaluar su desempeño, se visualizan las curvas de pérdida y se miden las métricas MSE y coeficiente de determinación (R^2), que permiten verificar la precisión del modelo.

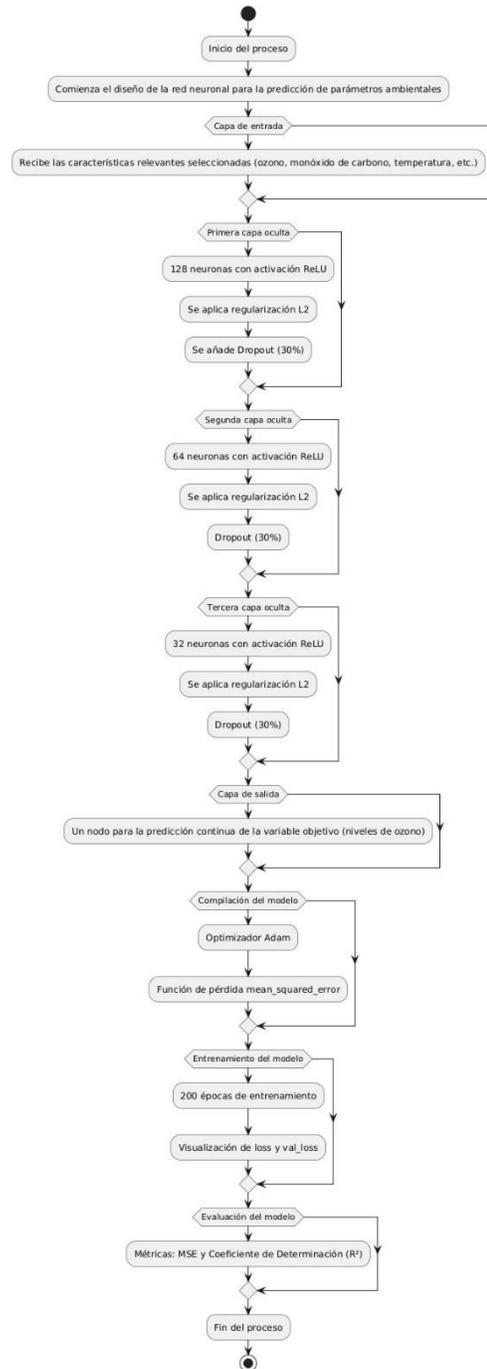


Figura 2. Diagrama de flujo del diseño y entrenamiento de la red MLP

RESULTADOS Y DISCUSIÓN

Correlación de las variables

La matriz de correlación presentada en la figura 3 muestra las relaciones entre distintas variables ambientales, destacando interacciones significativas y potencialmente útiles para el análisis y la modelización. La fuerte correlación negativa entre temperatura y humedad (-

0.969) indica que a medida que la temperatura aumenta, la humedad disminuye notablemente, lo que podría ser crucial para entender fenómenos climáticos y ambientales. Las correlaciones moderadas, tanto positivas como negativas, entre otras variables (ozono-humedad, ozono-propano, propano-CO) sugieren interacciones menos intensas pero relevantes, proporcionando información adicional sobre cómo estas variables pueden influir mutuamente. La casi inexistente correlación entre CO y temperatura (-0.018) indica una independencia entre estas dos variables en el contexto estudiado.

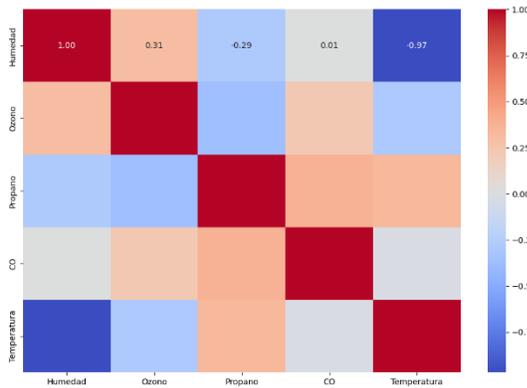


Figura 3. Matriz de correlación entre variables

Análisis Independiente de las Variables

En esta sección se realiza el análisis de la cada una de las variables para determinar la importancia de las características de la variable estudiada, el resultado del entrenamiento de cada una y un comparativo de los errores MSE y R^2 con los cuatro modelos aplicados (RL, DT, RF, NN).

Importancia de características de la Variable Humedad

Usando el modelo de RF se realizó un análisis de importancia de características, identificado a la temperatura, el propano y el ozono como las tres variables más influyentes para la predicción de la variable objetivo. La temperatura es la característica más relevante, lo que indica una relación alta con la variable objetivo, se puede atribuir a su impacto en las reacciones químicas y la volatilidad de los gases. La siguiente en importancia según el modelo, es el propano, probablemente por su vinculación con actividades industriales que generan este gas. El ozono es la tercera variable más relevante, aporta información adicional sobre la calidad del aire y las interacciones químicas en el entorno. Estas tres características seleccionadas se utilizarán para entrenar y evaluar nuestros modelos predictivos, asegurando predicciones precisas y confiables. La figura 4, muestra las características y su importancia para predecir la humedad.

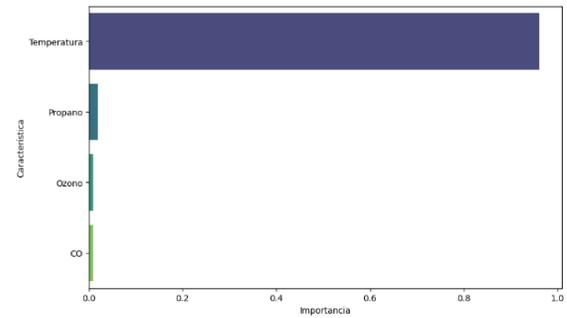


Figura 4. Importancia de características para predecir la humedad

Entrenamiento de la red neuronal para la variable Humedad

Durante el entrenamiento de la NN para predecir la humedad con datos de sensores, se observó una disminución significativa de la pérdida (loss) y la pérdida de validación (val_loss) a lo largo de las épocas. Inicialmente, la pérdida comenzó en 1596.39 y la pérdida de validación en 1542.45. A medida que avanzaron las épocas, la pérdida disminuyó a valores más bajos, alcanzando 13.44 en la época 93, mientras que la pérdida de validación fluctuó, llegando a un valor tan bajo como 1.39 en la época 80. Estos resultados indican que el modelo está aprendiendo de manera efectiva, la figura 5 muestra un el progreso de aprendizaje de las pérdidas por época tanto en el entrenamiento como en la validación.

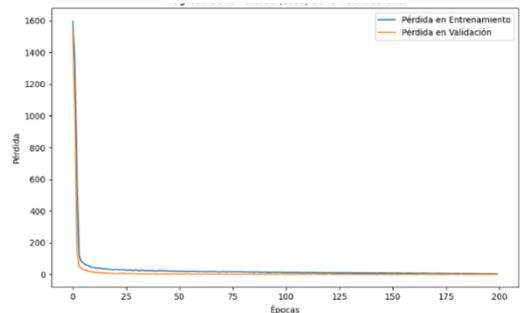


Figura 5. Progreso de Pérdidas de la NN para humedad

Las figuras 6, 7, 8 y 9 muestran un gráfico de dispersión comparando las predicciones de los cuatro modelos (LR, DT, NN y RF) entrenados para la predicción de la humedad, con los valores reales

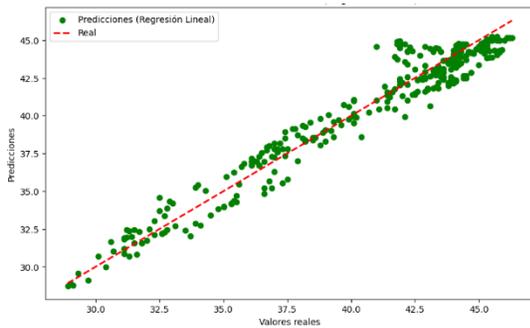


Figura 6. Predicciones vs Valores reales del modelo de RL para humedad

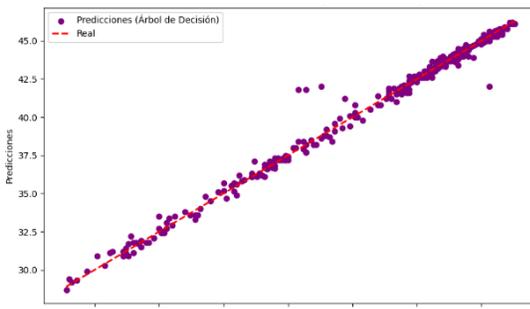


Figura 7. Predicciones vs Valores reales del modelo de DT para humedad

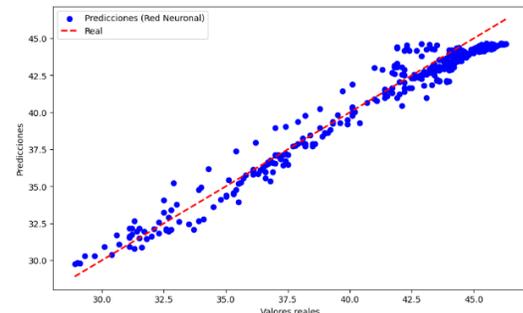


Figura 8. Predicciones vs Valores reales del modelo de NN para humedad

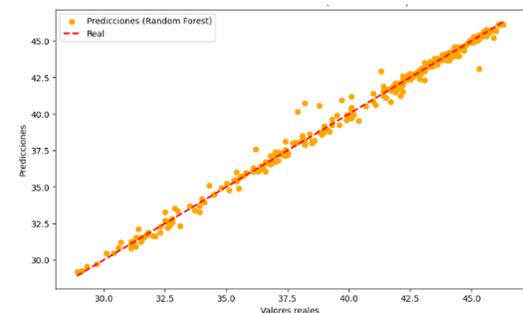


Figura 9. Predicciones vs Valores reales del modelo de RF para humedad

La tabla 3 resalta las características y el desempeño de cada modelo en términos de precisión y la presencia de desviaciones en sus predicciones.

Modelo/Figura	Precisión General	Desviaciones
Regresión Lineal Figura 5	Buen desempeño, alineándose bien con valores reales.	Algunas desviaciones, indicando errores de predicción más grandes en ciertos casos.
Árbol de Decisión Figura 6	Alto nivel de precisión.	Muy pocas desviaciones significativas.
Red Neuronal Figura 7	Buena precisión en general.	Más variabilidad y algunas predicciones menos precisas.
Random Forest Figura 8	Buen desempeño, preciso y confiable.	Algunas desviaciones, especialmente en los extremos.

Tabla 3. Comparación de las predicciones vs los valores reales de los modelos para la variable humedad

Aunque todos los modelos presentan un buen nivel de precisión, el DT y RF destacan por su exactitud y menor cantidad de desviaciones significativas. El modelo de NN, aunque preciso, muestra más variabilidad, y la LR, aunque efectiva, tiene algunas desviaciones más notables. Esto sugiere que, para la predicción de la humedad, los modelos de DT y RF podrían ser preferidos debido a su mayor precisión y consistencia.

Errores de los modelos entrenados de la Variable Humedad

Modelo	Error Cuadrático Medio (MSE)	Coefficiente de Determinación (R ²)
Regresión Lineal	1.1140	0.9476
Árbol de Decisión	0.2173	0.9898
Random Forest	0.1372	0.9936
Red Neuronal	0.8274	0.9611

Tabla 4. Resumen de los resultados de la evaluación de los modelos para humedad

La tabla 4 resume los resultados de la evaluación de los modelos en términos de precisión (MSE) y ajuste (R²), para predecir la humedad. El modelo de DT obtuvo el menor (MSE) con 0.2173, indicando la menor variabilidad entre las predicciones y los valores reales. Además, logró un alto (R²) de 0.9898, lo que sugiere que el 98.98% de la variabilidad en los datos puede ser explicada por el modelo. Le sigue el RF con un MSE de 0.1372 y un R² de 0.9936, demostrando una mejora adicional en precisión y ajuste. La NN, con un MSE de 0.8274 y un R² de 0.9611,

también mostró un buen desempeño general. Aunque su MSE es más alto que el de los modelos basados en árboles, su capacidad para modelar relaciones complejas sugiere que sigue siendo una opción viable para la predicción precisa de parámetros ambientales bajo condiciones variables y datos heterogéneos. Estos resultados resaltan la efectividad de los métodos de aprendizaje supervisado más avanzados en la tarea de predicción ambiental.

Importancia de características de la Variable Ozono

La matriz de importancia de características para predecir el ozono revela que el propano es la variable más influyente con una importancia del 40.3%, seguido por el CO con un 30.7%, indicando que estos gases tienen un fuerte impacto en la concentración de ozono, posiblemente debido a su participación en reacciones químicas atmosféricas. La humedad, con una importancia del 18.9%, también juega un papel significativo, afectando la formación y dispersión del ozono. La temperatura, aunque menos influyente con un 9.9%, contribuye a la predicción del ozono a través de su efecto en la velocidad de las reacciones químicas. En conjunto, estas variables clave son fundamentales para mejorar la precisión de los modelos predictivos de ozono., la figura 10, muestra las características y su importancia para predecir la humedad.

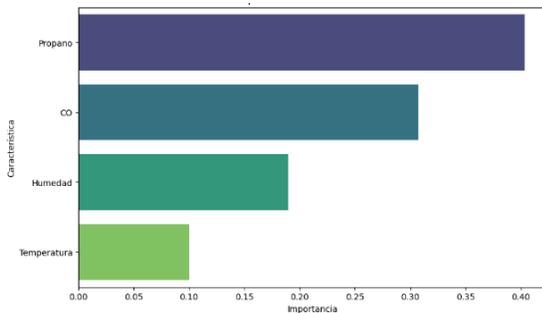


Figura 10. Importancia de características para predecir el ozono

Entrenamiento de la red neuronal para la variable ozono

La salida del entrenamiento del modelo muestra una disminución general de la pérdida tanto en el conjunto de entrenamiento como en el de validación a lo largo de las épocas, indicando que el modelo está aprendiendo a ajustar sus parámetros y a generalizar mejor. Sin embargo, se observan fluctuaciones en la pérdida de validación, lo que sugiere que, aunque el modelo está mejorando, hay ciertas variaciones en cómo se ajusta a los datos no vistos, la figura 11 muestra el desarrollo por etapas de esta NN.

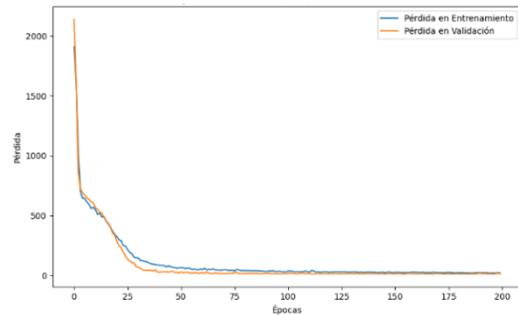


Figura 11. Progreso de Pérdidas de la NN para ozono

Las figuras 12, 13, 14 y 15 muestran un gráfico de dispersión comparando las predicciones de los cuatro modelos (LR, DT, NN y RF) entrenados para la predicción del ozono, con los valores reales

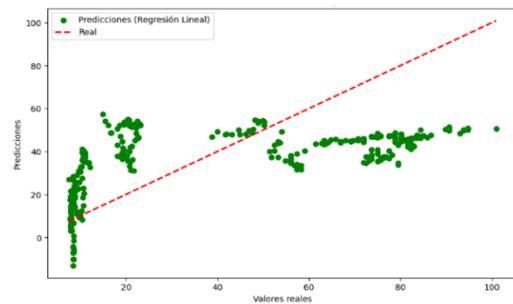


Figura 12. Predicciones vs Valores reales del modelo de LR para ozono

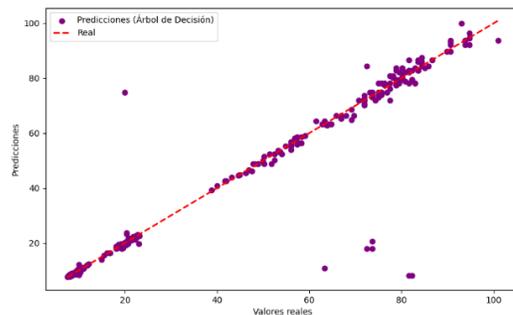


Figura 13. Predicciones vs Valores reales del modelo de DT para ozono

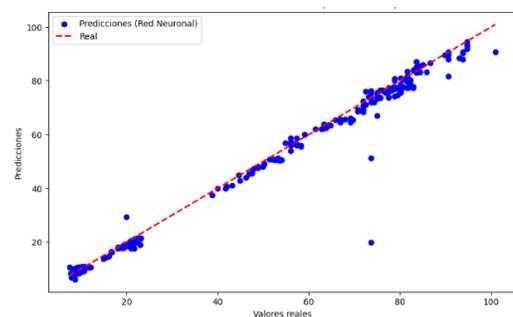


Figura 14. Predicciones vs Valores reales del modelo de NN para ozono

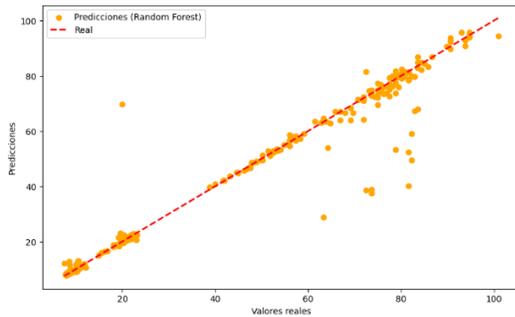


Figura 15. Predicciones vs Valores reales del modelo de RF para ozono

La tabla 5 resalta las características y el desempeño de cada modelo en términos de precisión y la presencia de desviaciones en sus predicciones, proporcionando una visión clara de sus fortalezas y debilidades.

Modelo/Figura	Precisión General	Desviaciones
Regresión Lineal Figura 12	Baja. La mayoría de los puntos se desvían de la línea ideal.	Significativas, especialmente en valores bajos y altos.
Árbol de Decisión Figura 13	Alta. La mayoría de los puntos están cerca de la línea roja.	Menores, aunque existen algunos puntos dispersos.
Red Neuronal Figura 14	Alta. La mayoría de los puntos se alinean bien con la línea roja.	Menores, con algunos puntos dispersos, pero menos frecuentes.
Random Forest Figura 15	Alta. La mayoría de los puntos están cerca de la línea roja.	Algunas desviaciones, especialmente en valores altos.

Tabla 5. Comparación de las predicciones vs los valores reales de los modelos para la variable ozono

Al comparar los modelos de LR, DT, NN y RF para predecir los niveles de ozono, se observa que los modelos de DT, NN y RF presentan una alta precisión general, con la mayoría de los puntos de predicción alineándose cerca de la línea ideal de referencia. En particular, el modelo de NN y RF muestran una excelente precisión, con menores desviaciones, aunque el RF presenta algunas inconsistencias en los valores más altos. En contraste, el modelo de LR tiene una baja precisión general, con desviaciones significativas, especialmente en los extremos de los valores predichos. Estos hallazgos sugieren que los modelos de DT, NN y RF son más adecuados para predecir los niveles de ozono, con el modelo de NN destacándose ligeramente por su capacidad para minimizar las desviaciones.

Errores de los modelos entrenados de la Variable Ozono

Modelo	Error Cuadrático Medio (MSE)	Coefficiente de Determinación (R ²)
Regresión Lineal	633.7956	0.3183
Árbol de Decisión	76.0096	0.9182
Random Forest	38.8838	0.9582
Red Neuronal	13.4554	0.9855

Tabla 6. Resumen de los resultados de la evaluación de los modelos para ozono

La tabla 6 muestra que, entre los cuatro modelos evaluados (LR, DT, NN, RF) para la predicción de los niveles de ozono, el modelo de NN es el más preciso, con el menor Error Cuadrático Medio (MSE) de 13.4554 y el mayor Coeficiente de Determinación (R²) de 0.9855, indicando que explica el 98.55% de la variabilidad de los datos. El modelo de RF sigue en precisión, con un MSE de 38.8838 y un R² de 0.9582. El DT también muestra buenos resultados con un MSE de 76.0096 y un R² de 0.9182. En contraste, el modelo de LR tiene el peor rendimiento, con un MSE significativamente más alto de 633.7956 y un R² de 0.3183, indicando una baja capacidad predictiva.

Importancia de características de la Variable Propano

Para predecir el nivel de Propano, el modelo considera principalmente el CO y el Ozono, con importancias de 0.570321 y 0.364791 respectivamente, indicando que estas características tienen un impacto significativo en las predicciones. La Humedad y la Temperatura, con importancias de 0.032478 y 0.032411, tienen una influencia mínima en comparación con el CO y el Ozono. Esto sugiere que el CO y el Ozono son los principales determinantes para la predicción del Propano, mientras que la Humedad y la Temperatura juegan un papel menor en el modelo, la figura 16 muestra un gráfico de observa la importancia de cada característica seleccionada para predecir el propano.

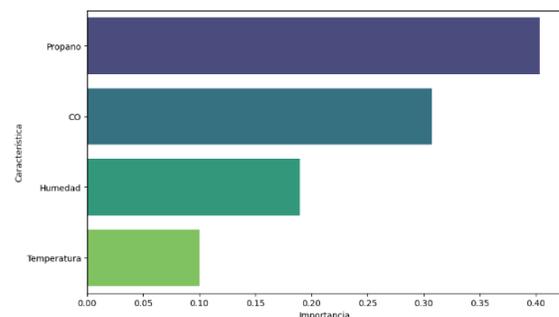


Figura 16. Importancia de características para predecir el propano

Entrenamiento de la red neuronal para la variable propano

El entrenamiento de la NN para predecir el propano muestra una tendencia positiva en la reducción de la pérdida de entrenamiento (loss) y la pérdida de validación (val_loss) a lo largo de las épocas. Desde la primera época hasta la 105, ambas métricas disminuyen consistentemente, indicando que el modelo está aprendiendo y mejorando su capacidad de predicción. En las primeras épocas, la pérdida es relativamente alta, pero a medida que avanza el entrenamiento, la pérdida de validación se estabiliza en valores bajos, sugiriendo que el modelo está generalizando bien y evitando el sobreajuste. La pérdida de validación alcanza valores muy bajos, como 0.0121 en la época 104, lo que sugiere que el modelo ha logrado un buen ajuste a los datos de entrenamiento y validación. La figura 17 muestra el desarrollo por etapas de esta NN.

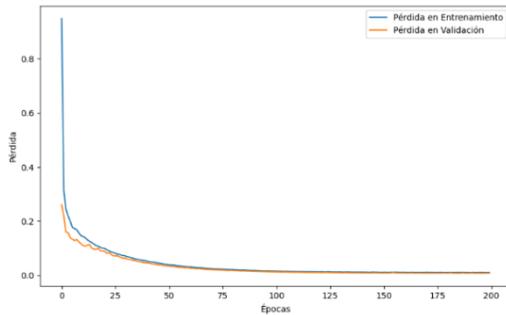


Figura 17. Progreso de Pérdidas de la NN para propano

Las figuras 18, 19, 20 y 21 muestran un gráfico de dispersión comparando las predicciones de los cuatro modelos (LR, DT, NN y RF) entrenados para la predicción del propano, con los valores reales

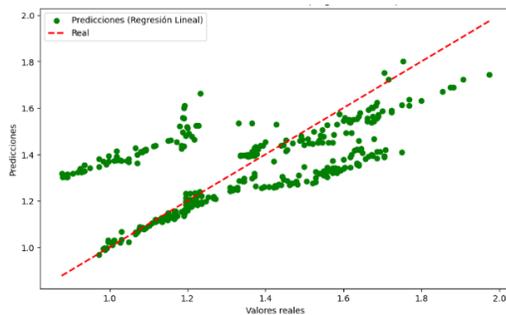


Figura 18. Predicciones vs Valores reales del modelo de RL para propano

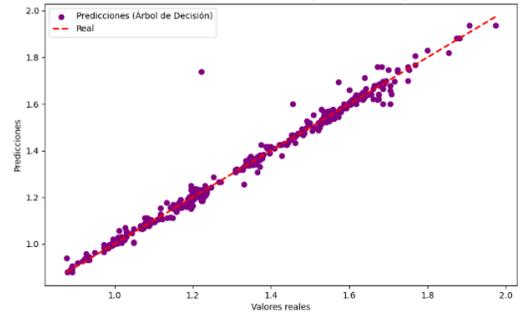


Figura 19. Predicciones vs Valores reales del modelo de DT para propano

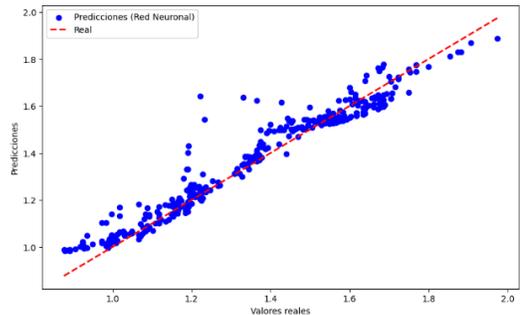


Figura 20. Predicciones vs Valores reales del modelo de NN para propano

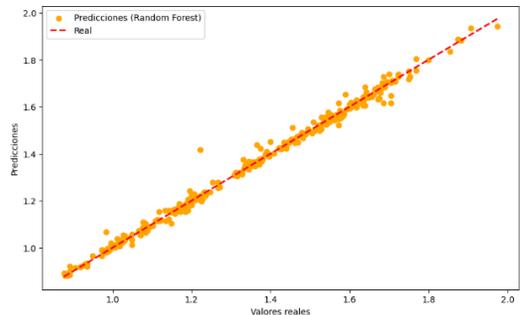


Figura 21. Predicciones vs Valores reales del modelo de RF para propano

Modelo/Figura	Precisión General	Desviaciones
Regresión Lineal Figura 18	Moderada	Alta dispersión alrededor de la línea diagonal
Árbol de Decisión Figura 19	Alta	Baja dispersión, puntos mayormente alineados con la línea diagonal
Red Neuronal Figura 20	Moderada-Alta	Dispersión intermedia, mejor que la LR, pero peor que DT
Random Forest Figura 21	Alta	Ligera dispersión, modelo realiza predicciones precisas en la mayoría de los casos

Tabla 7. Comparación de las predicciones vs los valores reales de los modelos para la variable propano

La tabla 7 muestra que el modelo de DT y RF presentan la mayor precisión general y las menores desviaciones en comparación con la LR y la NN. La LR tiene una precisión moderada y una alta dispersión, indicando un rendimiento inferior. La NN mejora respecto a la Regresión Lineal pero aún muestra una dispersión intermedia. En contraste, el DT y RF destacan por su alta precisión y baja dispersión, siendo RF ligeramente superior en términos de precisión general. En conclusión, para predecir los valores de propano, los modelos de DT y RF son los más recomendables debido a su precisión y consistencia en las predicciones.

Errores de los modelos entrenados de la Variable Propano

Modelo	Error Cuadrático Medio (MSE)	Coefficiente de Determinación R ²
Regresión Lineal	0.0375	0.3765
Árbol de Decisión	0.0013	0.9776
Random Forest	0.0004	0.9928
Red Neuronal	0.0040	0.9331

Tabla 8. Resumen de los resultados de la evaluación de los modelos para propano

La tabla 8 muestra una comparación entre diferentes modelos predictivos para el análisis de propano, evaluados mediante el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación R². Los resultados indican que el modelo de RF tiene el mejor desempeño, con el menor MSE (0.0004) y el mayor R² (0.9928), lo que sugiere una alta precisión en las predicciones y una excelente capacidad para explicar la variabilidad de los datos. El DT también presenta un buen rendimiento, con un MSE bajo (0.0013) y un R² elevado (0.9776). Por otro lado, la LR y la NN muestran desempeños inferiores, con MSE más altos (0.0375 y 0.0040, respectivamente) y R² menores (0.3765 y 0.9331). Los modelos complejos como RF y DT, superan a la LR y a la NN en términos de precisión y ajuste, lo que sugiere que pueden ser más adecuados para la predicción de propano en este contexto.

Importancia de características de la Variable CO

La figura 22, muestra un gráfico de observa la importancia de características del modelo de RF de cada característica seleccionada para predecir el CO. revela que el propano. La variable más influyente, contribuyendo con un 63.22% a las predicciones, seguido por el ozono con un 30.83%. La temperatura y la humedad tienen una influencia significativamente menor, con importancias del 3.77% y 2.18%, respectivamente. Esto indica que las variaciones

en los niveles de propano y ozono son los factores más determinantes para la precisión del modelo, mientras que la temperatura y la humedad apenas afectan las predicciones.

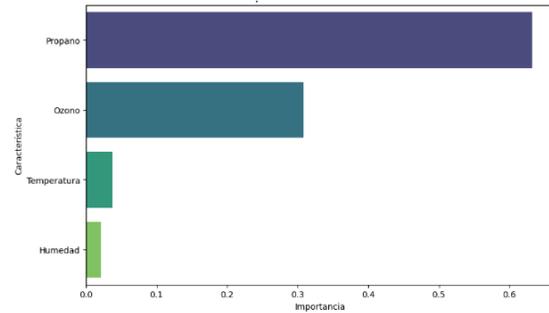


Figura 22. Importancia de características para predecir el CO

Entrenamiento de la red neuronal para la variable CO

El entrenamiento de la NN muestra una progresiva reducción en la función de pérdida, indicando que el modelo está mejorando su capacidad para predecir en cada época. Desde el inicio, donde la pérdida era significativamente alta, el modelo muestra una notable disminución en la pérdida de entrenamiento y de validación, alcanzando un mínimo en la época 147, donde la pérdida de validación se estabiliza en torno a 2.6752. Esto sugiere que el modelo ha aprendido a generalizar mejor a los datos de validación con el tiempo. La figura 23 muestra el desarrollo por etapas de esta NN.

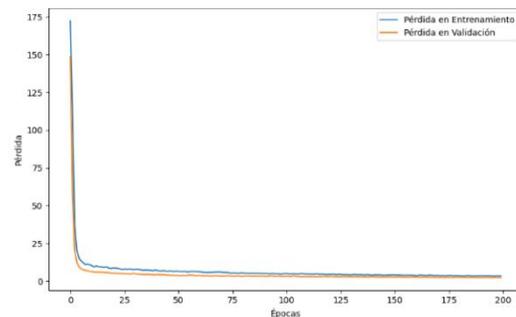


Figura 23. Progreso de Pérdidas de la NN para CO

Las figuras 24, 25, 26 y 27 muestran un gráfico de dispersión comparando las predicciones de los cuatro modelos (LR, DT, NN, RF) entrenados para la predicción del CO, con los valores reales

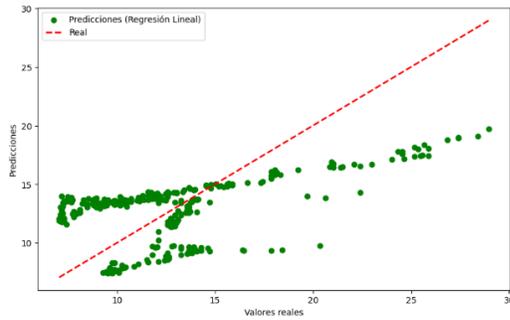


Figura 24. Predicciones vs Valores reales del modelo de RL para CO

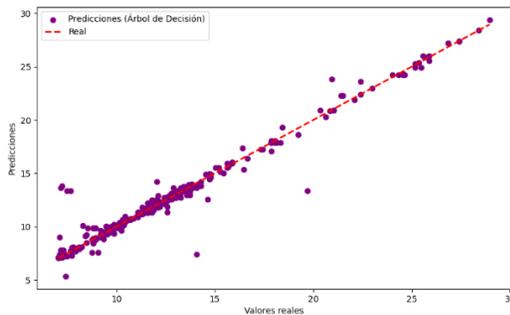


Figura 25. Predicciones vs Valores reales del modelo de DT para CO

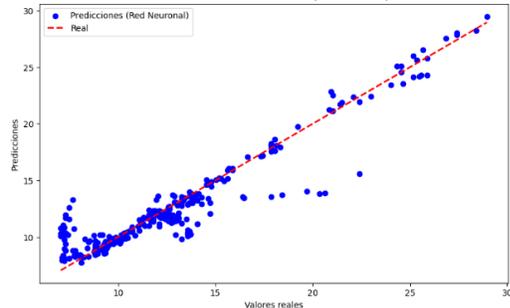


Figura 26. Predicciones vs Valores reales del modelo de NN para CO

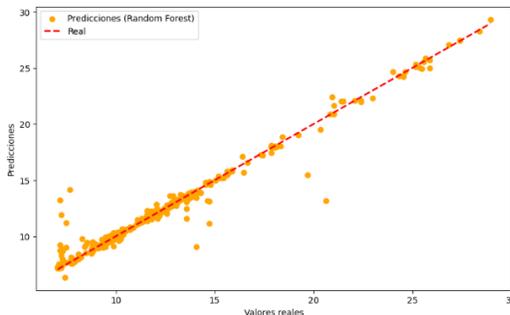


Figura 27. Predicciones vs Valores reales del modelo de RF para CO

Modelo/Figura	Precisión General	Desviaciones
Árbol de Decisión Figura 24	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Aprendizaje por Refuerzo (RL) Figura 25	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Red Neuronal Figura 26	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Random Forest Figura 27	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.

Tabla 9. Comparación de las predicciones vs los valores reales de los modelos para la variable CO

La tabla 9 muestra que todos los modelos evaluados (DT, LR, NN y RF) demuestran una capacidad similar para predecir los valores de CO, con una precisión general que sugiere que todos son útiles para la tarea. Sin embargo, cada modelo presenta un error moderado, evidenciado por una ligera dispersión de los puntos alrededor de la línea diagonal en las gráficas de dispersión. Esto indica que, aunque las predicciones son generalmente precisas, ningún modelo es perfecto y todos presentan un margen de error en sus estimaciones.

Errores de los modelos entrenados de la Variable CO

Modelo	Error Cuadrático Medio (MSE)	Coefficiente de Determinación R ²
Regresión Lineal	14.2036	0.2639
Árbol de Decisión	0.8722	0.9548
Random Forest	0.8020	0.9584
Red Neuronal	2.1607	0.8880

Tabla 10. Resumen de los resultados de la evaluación de los modelos para CO

Al comparar los resultados de la tabla 10 donde se muestran el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación R² con los datos de precisión general y desviaciones obtenidos para cada modelo, se puede observar que el RF es el modelo que ofrece el mejor desempeño general, con el menor MSE (0.8020) y el mayor R² (0.9584). Esto indica que es el

modelo más preciso y con mayor capacidad para explicar la variabilidad en los valores reales de CO. El DT, aunque tiene un MSE (0.8722) ligeramente mayor que el RF, sigue siendo un modelo sólido con un R² muy alto (0.9548), lo que sugiere una buena precisión general y desviaciones relativamente bajas. La NN, a pesar de tener un R² bastante bueno (0.8880), presenta un MSE considerablemente mayor (2.1607), lo que indica que su precisión general es menor en comparación con los modelos basados en árboles. Por último, la LR muestra el peor desempeño en ambas métricas, con el mayor MSE (14.2036) y el R² más bajo (0.2639), lo que revela una precisión muy baja y una capacidad limitada para explicar la variabilidad de los datos.

Importancia de características de la Variable Temperatura

La Figura 28 presenta una gráfica de Importancia de Características para la predicción de la temperatura. Los datos muestran que la Humedad es, con diferencia, la variable más influyente con una importancia de 0.969379, lo que indica que es el principal factor que afecta la predicción de la temperatura en el modelo. En contraste, las otras características tienen una influencia mucho menor: Propano con una importancia de 0.016481, Ozono con 0.008229 y CO con 0.005912. Estas bajas importancias sugieren que, aunque están incluidas en el modelo, su impacto en la predicción de la temperatura es prácticamente insignificante comparado con la humedad. Los resultados de la Figura 27 destacan la crucial relevancia de la humedad en la precisión del modelo de predicción de la temperatura.

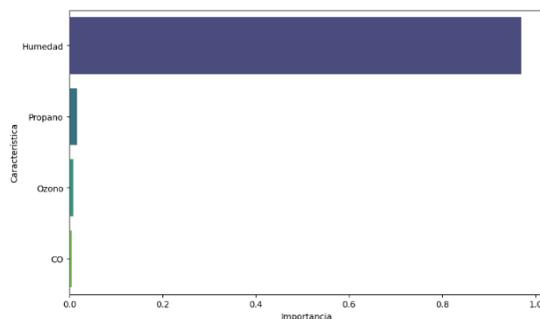


Figura 28. Importancia de características para predecir la Temperatura

Entrenamiento de la red neuronal para la variable Temperatura

La figura 29 muestra los resultados de entrenamiento de un modelo durante 200 épocas, registrando la pérdida del modelo y la pérdida de validación. En las primeras épocas, la pérdida del modelo disminuye drásticamente, luego se estabiliza hasta llegar a valores mucho más bajos, como 4.3426 en la época 116. La pérdida de validación también muestra una tendencia a la baja, comenzando en 536.1054

y alcanzando valores mínimos como 0.3670 en la época 115. Sin embargo, hay fluctuaciones ocasionales en ambas métricas, lo cual es normal durante el proceso de entrenamiento. Estos resultados indican que el modelo está aprendiendo y mejorando su rendimiento en la tarea asignada, aunque algunas épocas pueden presentar sobreajuste o fluctuaciones debidas a la naturaleza de los datos.

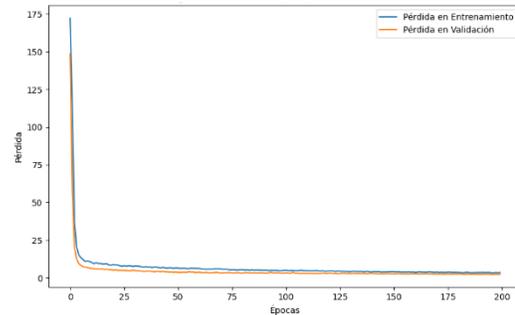


Figura 29. Progreso de Pérdidas de la NN para Temperatura

Las figuras 30, 31, 32 y 33 muestran un gráfico de dispersión comparando las predicciones de los cuatro modelos (LR, DT, NN, RF) entrenados para la predicción del CO, con los valores reales.

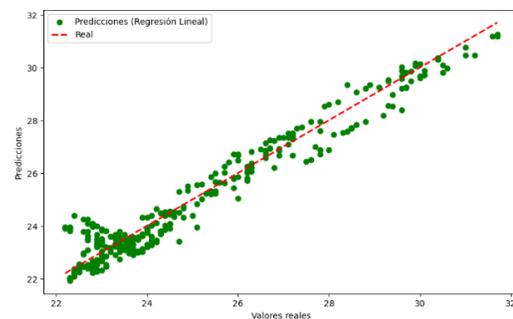


Figura 30. Predicciones vs Valores reales del modelo de RL para Temperatura

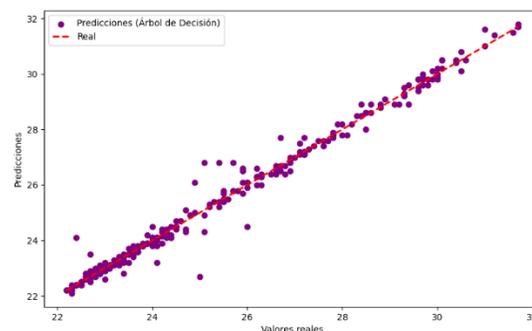


Figura 31. Predicciones vs Valores reales del modelo de DT para Temperatura

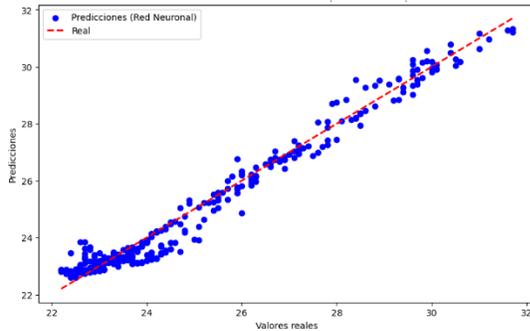


Figura 32. Predicciones vs Valores reales del modelo de red neuronal para Temperatura

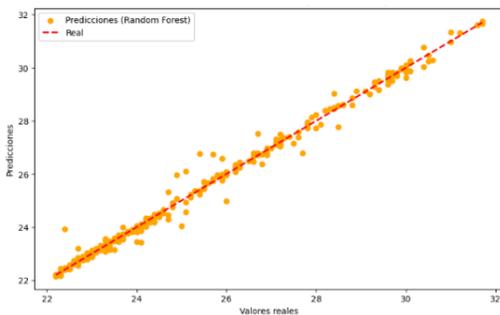


Figura 33. Predicciones vs Valores reales del modelo de RF para Temperatura

Modelo/Figura	Precisión General	Desviaciones
Regresión Lineal (RL) Figura 30	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Árbol de Decisión Figura 31	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Red Neuronal Figura 32	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.
Random Forest Figura 33	El modelo realiza predicciones precisas para la mayoría de los casos.	Error moderado, con ligera dispersión de los puntos alrededor de la línea diagonal.

Tabla 11. Comparación de las predicciones vs los valores reales de los modelos para la variable Temperatura

La tabla 11 compara los modelos para la predicción de la temperatura, muestra que todos los modelos (LR, DT, RF Y NN) realizan predicciones precisas para la mayoría de los casos. Sin embargo, presentan un error moderado con una ligera dispersión de los puntos alrededor de la línea

diagonal en las gráficas de dispersión. Esto sugiere que, aunque los modelos son generalmente fiables, aún pueden cometer errores. En términos de precisión general y desviaciones, no se observa una diferencia significativa entre los modelos, lo que indica que todos son igualmente competentes para esta tarea específica.

Errores de los modelos entrenados de la Variable Temperatura

Modelo	Error Cuadrático Medio (MSE)	Coefficiente de Determinación R ²
Regresión Lineal	0.3152	0.9489
Árbol de Decisión	0.0860	0.9861
Random Forest	0.0529	0.9914
Red Neuronal	0.1630	0.9736

Tabla 12. Resumen de los resultados de la evaluación de los modelos para Temperatura

La tabla 12 proporciona una visión detallada de la precisión de diversos modelos de predicción, reflejando dos métricas clave: el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación R². La LR presenta un MSE de 0.3152 y un R² de 0.9489, lo que indica un ajuste aceptable pero menos preciso en comparación con los otros modelos. El DT muestra una mejora significativa con un MSE de 0.0860 y un R² de 0.9861, sugiriendo una mayor precisión en las predicciones. El RF destaca con el MSE más bajo de 0.0529 y el R² más alto de 0.9914, lo que demuestra un rendimiento superior en términos de exactitud y ajuste al conjunto de datos. La NN, aunque robusta, tiene un MSE de 0.1630 y un R² de 0.9736, que, aunque buenos, son inferiores a los del RF y el DT.

En comparación con los resultados de la tabla 10, donde todos los modelos mostraron un error moderado y una dispersión ligera alrededor de la línea diagonal, la tabla actual revela que el RF Y DT ofrecen un ajuste mucho más preciso y explicativo. La NN y la LR, aunque útiles, muestran una menor precisión y capacidad de ajuste en comparación con estos modelos de árbol, subrayando la superioridad del RF en esta evaluación.

CONCLUSIONES

El análisis comparativo de los modelos de predicción de parámetros ambientales ha proporcionado valiosas observaciones sobre su desempeño y aplicabilidad. A través de la evaluación de modelos como LR, DT, NN Y RF, se han identificado las fortalezas y debilidades de cada enfoque para predecir variables ambientales específicas. Los hallazgos de este estudio son esenciales para guiar la selección de modelos en aplicaciones prácticas,

optimizando la precisión y eficiencia en la predicción de condiciones ambientales.

Precisión de Modelos: Árbol de Decisión y Random Forest demostraron ser los más precisos para la predicción de humedad, con errores cuadráticos medios (MSE) de 0.2173 y 0.1372 respectivamente, y altos coeficientes de determinación (R^2) de 0.9898 y 0.9936. Estos modelos superan significativamente a la LR y la NN en términos de precisión y consistencia para esta variable.

Para la predicción de ozono, la **Red Neuronal** se destacó con un MSE de 13.4554 y un R^2 de 0.9855, mostrando una excelente capacidad para manejar relaciones complejas en datos ambientales.

Random Forest también tuvo el mejor desempeño general para la predicción de monóxido de carbono (CO), con el menor MSE y el mayor R^2 , indicando alta precisión y capacidad para explicar la variabilidad en los datos.

Eficiencia y Adaptabilidad: La **Red Neuronal**, aunque mostró una mayor variabilidad en sus predicciones de humedad, fue extremadamente precisa para la predicción de niveles de ozono. Esto subraya su capacidad para manejar relaciones complejas en datos ambientales, adaptándose bien a las variaciones en los datos.

Los modelos de **Árbol de Decisión** y **Random Forest** mostraron consistencia y precisión en múltiples variables, indicando su eficiencia y adaptabilidad en diferentes contextos de datos.

Aplicabilidad Práctica: Los modelos basados en árboles, como el **Árbol de Decisión** y **Random Forest**, son preferibles para aplicaciones donde la precisión y la consistencia son críticas, como en la gestión de recursos hídricos y la agricultura. Su alta precisión y consistencia los hacen ideales para predicciones donde los errores pueden tener un impacto significativo.

Las **Redes Neuronales** son más adecuadas para la predicción de parámetros con una mayor variabilidad, como la calidad del aire, debido a su capacidad para manejar relaciones no lineales y complejas en los datos.

Impacto en la Gestión Ambiental: La implementación de modelos de **Árbol de Decisión** y **Random Forest** puede mejorar significativamente la capacidad de monitoreo y respuesta a condiciones ambientales cambiantes, contribuyendo a la toma de decisiones informada en áreas como la salud pública y la gestión urbana.

La **Red Neuronal**, con su alta precisión en la predicción de ozono, puede ser particularmente útil en la gestión de la calidad del aire, permitiendo respuestas rápidas y precisas a los cambios en los niveles de contaminantes.

El estudio sugiere que los modelos de DT Y RF son preferibles para la mayoría de las variables ambientales evaluadas debido a su alta precisión y consistencia. La NN también es efectiva, especialmente en la predicción de

ozono, aunque puede presentar mayor variabilidad en otras variables. La LR generalmente mostró el peor desempeño y puede no ser adecuada para predicciones precisas en contextos ambientales complejos. Estos hallazgos resaltan la importancia de seleccionar el modelo adecuado según el parámetro ambiental específico y las condiciones de los datos, optimizando así la precisión y efectividad de las predicciones.

CONTRIBUCIÓN DE LOS AUTORES

Autor Principal:

Creación de los modelos predictivos.

Análisis de datos.

Procesamiento de datos.

Redacción del manuscrito.

Autor 2:

Revisión crítica del contenido.

Recolección de datos.

Apoyo en la interpretación de los resultados y la formulación de las conclusiones.

REFERENCIAS BIBLIOGRÁFICAS

- Alotaibi, E., Omar, M., Arab, M. G., & Tahmaz, A. (2022). Prediction of Fine-Grained Soils Shrinkage Limits Using Artificial Neural Networks. *2022 Advances in Science and Engineering Technology International Conferences, ASET 2022*. <https://doi.org/10.1109/ASET53988.2022.9734806>
- Alwan, H. K., Al-Sultany, G. A., & Hussein, S. M. (2023). Air Pollution Prediction Using Machine Learning and Neural Network. *5th International Conference on Information Technology, Applied Mathematics and Statistics, ICITAMS 2023*, 231–234. <https://doi.org/10.1109/ICITAMS57610.2023.10525306>
- Baranitharan, B., Sivapragasam, C., & Rajesh, K. (2021). Energy Demand Prediction Considering Weather Change Effects using Artificial Neural Network. *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, 916–921. <https://doi.org/10.1109/ICAC3N53548.2021.9725452>
- Cen, H., Li, J., Liu, Z., Nie, J., & Cai, Q. (2023). Temperature prediction model of sheep barn in winter based on Neural Network. *2023 24th International Conference on Electronic Packaging Technology, ICEPT 2023*. <https://doi.org/10.1109/ICEPT59018.2023.10492107>
- Chen, J., Wei, X., Liu, Y., Liu, Z., Ma, Y., Zhang, M., Guo, Y., Bao, Z., & Zhang, H. (2023). Prediction of Water Quality of Xincheng Bridge in Lanzhou Based on CNN-LSTM Model. *Proceedings - 2023 3rd International Signal Processing*,

- Communications and Engineering Management Conference, ISPCEM 2023*, 59–63. <https://doi.org/10.1109/ISPCEM60569.2023.00018>
- Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. In *Applied Sciences (Switzerland)* (Vol. 10, Issue 17). <https://doi.org/10.3390/app10175776>
- Ding, S., Yang, H., Wang, Z., Song, G., Peng, Y., & Peng, X. (2018). Dynamic Prediction of the Silicon Content in the Blast Furnace using LSTM-RNN-Based Models. *Proceedings of 2018 International Computers, Signals and Systems Conference, ICOMSSC 2018*, 491–495. <https://doi.org/10.1109/ICOMSSC45026.2018.8941807>
- Frolov, S. V., Potlov, A. Y., Korobov, A. A., & Savinova, K. S. (2023). Neural Network Control of Environmental Parameters in Neonatal Incubators. *Proceedings of 2023 4th International Conference on Neural Networks and Neurotechnologies, NeuroNT 2023*, 21–24. <https://doi.org/10.1109/NEURONT58640.2023.10175837>
- Huu, K. D. (2024). Research on Ship Weather Routing Method Based on Dijkstra Algorithm and Neural Network. *Proceedings - 2024 International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM 2024*, 739–744. <https://doi.org/10.1109/ICIEAM60818.2024.10553838>
- Manocha, A., Sood, S. K., & Bhatia, M. (2024). IoT-Dew Computing-Inspired Real-Time Monitoring of Indoor Environment for Irregular Health Prediction. *IEEE Transactions on Engineering Management*, 71, 1669–1682. <https://doi.org/10.1109/TEM.2023.3338458>
- Pandya, M., Dave, V., & Ghosh, R. (2020). Artificial neural network (ANN) based soil electrical conductivity (SEC) prediction. *2020 7th International Conference on Signal Processing and Integrated Networks, SPIN 2020*, 581–586. <https://doi.org/10.1109/SPIN48934.2020.9071257>
- Popli, N., Davoodi, E., Capitanescu, F., & Wehenkel, L. (2024). Machine Learning Based Binding Contingency Pre-Selection for AC-PSCOPF Calculations. *IEEE Transactions on Power Systems*, 39(2), 4751–4754. <https://doi.org/10.1109/TPWRS.2023.3338971>
- Rao, D. V. S. K. K., Prusty, B. R., & Myneni, H. (2022). Bright Sunshine Duration Index-Based Prediction of Solar PV Power Using ANN Approach. *2022 International Conference on Intelligent Controller and Computing for Smart Power, ICICCSPP 2022*. <https://doi.org/10.1109/ICICCSPP53532.2022.9862452>
- Sathyaseelan, K., & Sarathambekai, S. (2021). Machine Learning based Prediction Model for Health Care Sector - A Survey. *3rd IEEE International Virtual Conference on Innovations in Power and Advanced Computing Technologies, i-PACT 2021*. <https://doi.org/10.1109/I-PACT52855.2021.9696646>
- Shu, B., Zhang, W., Chen, Y., Sun, J., & Wang, C. X. (2024). Path Loss Prediction in Evaporation Ducts Based on Deep Neural Network. *IEEE Antennas and Wireless Propagation Letters*, 23(2), 798–802. <https://doi.org/10.1109/LAWP.2023.3336469>
- Shukla, A., Sharma, M., Tiwari, K., Vani, V. D., Kumar, N., & Pooja. (2023). Predicting Rainfall Using an Artificial Neural Network-Based Model. *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, 2700–2704. <https://doi.org/10.1109/IC3I59117.2023.10397714>
- Suleman, M. A. R., & Shridevi, S. (2022). Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model. *IEEE Access*, 10, 82456–82468. <https://doi.org/10.1109/ACCESS.2022.3196381>
- Wu, L., He, D., Guan, K., Ai, B., Briso-Rodriguez, C., Shui, T., Liu, C., Zhu, L., & Shen, X. (2020). Received Power Prediction for Suburban Environment based on Neural Network. *International Conference on Information Networking, 2020-January*, 35–39. <https://doi.org/10.1109/ICOIN48656.2020.9016532>
- Yang, H., Yu, H., Zheng, K., Hu, J., Tao, T., & Zhang, Q. (2023). Hyperspectral Image Classification Based on Interactive Transformer and CNN With Multilevel Feature Fusion Network. *IEEE Geoscience and Remote Sensing Letters*, 20. <https://doi.org/10.1109/LGRS.2023.3303008>
- Yang, J., Chen, H., Lin, S., Chen, L., & Chen, Y. (2022). Prediction of temperature change with multi-dimensional environmental characteristic based on CNN-LSTM-ATTENTION model. *IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2022-June*, 1024–1029. <https://doi.org/10.1109/ITAIC54216.2022.9836837>