



Predicción de NOx en Condiciones Reales Mediante Modelos de Deep Learning

Electiva Deep Learning – Facultad de Ingeniería

Sergio Andrés Zapata Gomez
Universidad de Antioquia

4 de diciembre de 2025

1. Introducción y contexto

El trabajo desarrollado en este proyecto se enmarca en el estudio de emisiones vehiculares bajo condiciones reales de conducción. El foco se sitúa en la predicción de las concentraciones instantáneas de óxidos de nitrógeno (NOx) a partir de registros telemétricos obtenidos en tres vehículos distintos. Estos registros incluyen tanto variables de operación del motor como información ambiental y parámetros asociados al flujo de gases de escape. La existencia de esta infraestructura de sensórica abre la puerta a construir modelos que relacionen directamente las variables de uso del vehículo con el comportamiento del contaminante, sin necesidad de disponer de equipos de medición dedicados en todo momento.

Desde el punto de vista metodológico, el proyecto tuvo dos objetivos complementarios. Por un lado, organizar un flujo de trabajo reproducible en Google Colab, estructurado en cuadernos numerados que pudieran ejecutarse sin intervención manual adicional. Por otro lado, evaluar comparativamente varias arquitecturas de aprendizaje profundo: modelos densos (MLP), redes convolucionales unidimensionales (CNN 1D) y redes de tipo LSTM, además de dos referencias básicas: un predictor de media y una regresión lineal. Cada modelo se entrenó y validó sobre el mismo conjunto de datos, utilizando siempre la misma división temporal y el mismo preprocesamiento, para que las diferencias observadas pudieran atribuirse a las arquitecturas y no a variaciones en los datos.

2. Dataset, preprocesamiento y organización de los notebooks

Los datos iniciales se presentaron en tres archivos CSV correspondientes a los vehículos Etios, Figo y RRV. Cada archivo contiene una serie temporal multivariada donde las filas representan instantes de tiempo y las columnas corresponden a variables físicas: velocidad del vehículo, revoluciones por minuto, cargas del motor, presiones y temperaturas en distintos puntos del sistema, humedad y presión ambiente, así como concentraciones húmedas (*wet*) y masas de varios gases en el escape. La variable objetivo seleccionada para todo el proyecto fue `NOx_wet_conc`, que representa la concentración instantánea de NOx en condiciones húmedas.

La primera fase del trabajo consistió en inspeccionar la estructura de los archivos y unificarlos en un solo conjunto. Se verificó la consistencia de las columnas y se descartaron únicamente aquellas variables con carácter identificador (por ejemplo, nombres de conductor o etiquetas de ruta) que no aportaban información física directa y que podrían inducir al modelo a aprender patrones espurios si se mantenían. Una vez unificados los tres archivos, se ordenó el conjunto de

datos por la marca temporal disponible, de modo que la secuencia respetara el orden real de los eventos registrados.

El preprocesamiento se centró en dos aspectos: la normalización y la división temporal del conjunto. Para homogeneizar la escala entre variables con unidades diferentes se utilizó un escalado estándar basado en la media y desviación estándar del subconjunto de entrenamiento. De forma explícita, para cada característica x se aplicó

$$\tilde{x} = \frac{x - \mu_{\text{train}}}{\sigma_{\text{train}}}, \quad (1)$$

donde μ_{train} y σ_{train} se calcularon sólo con datos de entrenamiento. Esta decisión evita que la información del conjunto de validación o de prueba influya en el proceso de normalización. Tras el escalado, el conjunto se dividió en 70 % para entrenamiento, 15 % para validación y 15 % para prueba, conservando el orden temporal para respetar la naturaleza secuencial de los datos.

Con el fin de simplificar la reutilización de estos datos en distintos cuadernos, se generó un archivo comprimido `splits_escalados.npz` que almacena, ya normalizados, los arreglos `X_train`, `X_val`, `X_test` y sus correspondientes etiquetas `y_train`, `y_val`, `y_test`. Este archivo se ubicó en Google Drive con acceso público y se cargó en los notebooks mediante una URL directa. Con ello, cualquier usuario puede reproducir los experimentos sin necesidad de montar manualmente su unidad de Drive ni ejecutar el preprocesamiento desde cero.

La estructura global del proyecto se organizó en siete notebooks. El primero se dedicó a la exploración inicial del dataset y a la verificación de la variable objetivo. El segundo realizó todo el preprocesamiento descrito y produjo el archivo comprimido. El tercero implementó los modelos baseline y el MLP, mientras que el cuarto y quinto se centraron en la CNN 1D y la LSTM respectivamente. El sexto se reservó para una posible arquitectura híbrida CNN-LSTM, que no se utilizó para el análisis final, y el séptimo cuaderno consolidó las métricas y generó las gráficas comparativas.

3. Modelos y procedimiento de entrenamiento

Los modelos se entrenaron utilizando siempre la misma variable objetivo y las mismas entradas escaladas. Como punto de partida se utilizó un modelo baseline que predice la media de las observaciones de entrenamiento. Aunque su desempeño es pobre, ofrece una referencia mínima: cualquier modelo útil debe mejorar de manera clara este error. A continuación se ajustó una regresión lineal ordinaria, con la forma

$$\hat{y} = \beta^\top X, \quad (2)$$

donde β es el vector de parámetros ajustado mediante mínimos cuadrados. Este modelo ya incorpora relaciones lineales entre las variables, y de hecho mostró un rendimiento considerablemente mejor que el baseline de media.

El modelo de perceptrón multicapa (MLP) se definió con una entrada de 46 características, dos capas densas intermedias de 64 y 32 neuronas con activación `ReLU`, y una salida escalar para la concentración de NOx. El entrenamiento se realizó utilizando la pérdida de error cuadrático medio (MSE) y el optimizador Adam, con un número de épocas moderado que permitió observar la evolución tanto del error de entrenamiento como del de validación. Para los modelos secuenciales (CNN y LSTM) se construyeron ventanas temporales de longitud $W = 30$. Cada muestra de entrada para estos modelos tiene tamaño 30×46 , donde las 30 filas representan pasos consecutivos en el tiempo y las columnas corresponden a las mismas variables físicas usadas en el MLP.

La CNN 1D se diseñó con dos capas convolucionales con 32 filtros y tamaño de kernel 3, seguidas de una operación de aplanado y una capa densa previa a la salida. Esta arquitectura trata de capturar patrones locales en la ventana temporal, por ejemplo cambios rápidos en la

combinación de variables que preceden a picos en la concentración de NOx. La LSTM, por su parte, se implementó con 64 unidades recurrentes y una capa densa intermedia de 32 neuronas. El interés era estudiar si la estructura recurrente añadía capacidad para manejar dependencias de mayor alcance en la secuencia de medidas.

Todas las arquitecturas se evaluaron con las métricas MAE, RMSE y R^2 . Para la comparación de curvas de entrenamiento y validación se empleó MSE, puesto que coincide con la función de pérdida utilizada durante el ajuste de los pesos.

4. Análisis de las curvas de entrenamiento

Las curvas de entrenamiento y validación ofrecen una vista directa del comportamiento de cada modelo durante el ajuste. En el caso del MLP, la Figura 1 muestra cómo la pérdida de entrenamiento decrece de forma suave a lo largo de las épocas, mientras que la pérdida de validación desciende con rapidez en las primeras iteraciones y luego se estabiliza en un rango bajo, con ligeras oscilaciones. La separación entre ambas curvas es pequeña, lo que indica que el modelo logra ajustarse bien a los datos de entrenamiento sin deteriorar su desempeño sobre el conjunto de validación. La escala logarítmica del eje vertical revela que el MSE se sitúa, tras unas pocas épocas, por debajo de 10^{-1} . Este comportamiento consistente se traduce en las métricas finales, donde el MLP alcanza un error absoluto medio cercano a 0.26.

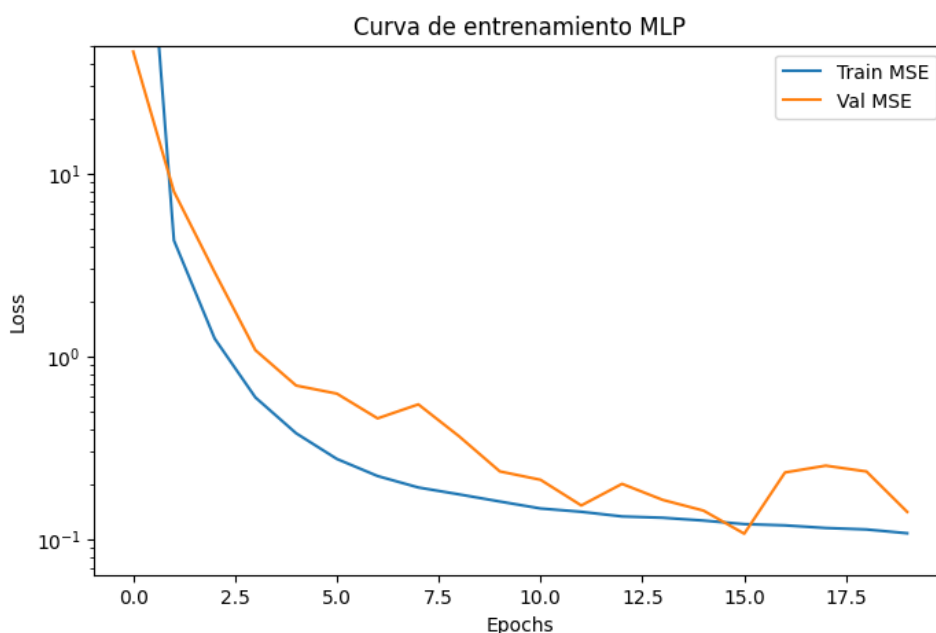


Figura 1: Curva de entrenamiento del MLP (MSE en escala logarítmica).

La Figura 2 presenta el comportamiento de la CNN 1D. Al inicio se observa una caída pronunciada del MSE tanto en entrenamiento como en validación. La curva de entrenamiento desciende de forma monótonica, mientras que la de validación presenta una ligera meseta alrededor de la época 5 y algunas oscilaciones posteriores. Aun así, ambas siguen trayectorias razonablemente cercanas, lo que sugiere que la CNN también mantiene una generalización aceptable. La pérdida de validación se estabiliza en un rango de MSE del orden de unidades, mayor que en el MLP, pero claramente inferior a los modelos lineales. El descenso inicial rápido indica que la arquitectura captura patrones locales útiles en las primeras épocas, y las oscilaciones posteriores pueden asociarse a ajustes finos de los filtros y pesos densos.

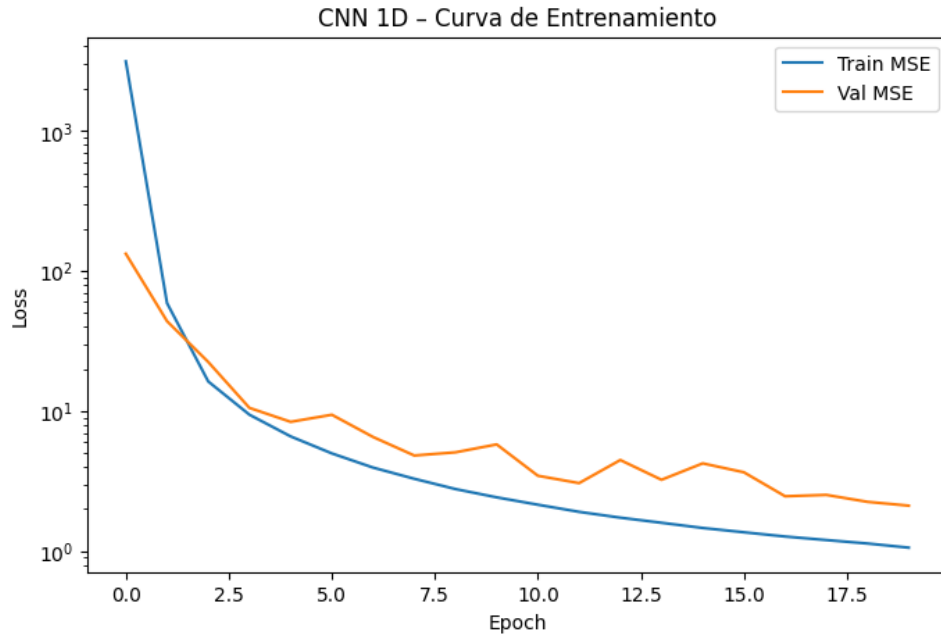


Figura 2: Curva de entrenamiento de la CNN 1D (MSE en escala logarítmica).

El caso de la LSTM es diferente. En la Figura 3 se aprecia un descenso inicial del error de entrenamiento similar al de las otras arquitecturas, pero la curva de validación no mantiene la misma tendencia. Tras una reducción durante las primeras épocas, el MSE de validación comienza a aumentar de forma progresiva, mientras que el de entrenamiento sigue disminuyendo. La brecha entre ambas curvas se ensancha a medida que avanzan las épocas. Este patrón es característico de un sobreajuste persistente: la red recurrente continúa ajustándose a detalles de las secuencias de entrenamiento que no se reproducen con la misma consistencia en los datos de validación. La escala logarítmica del eje vertical ayuda a resaltar este comportamiento, ya que el error de validación pasa de valores del orden de cientos a superar los miles.

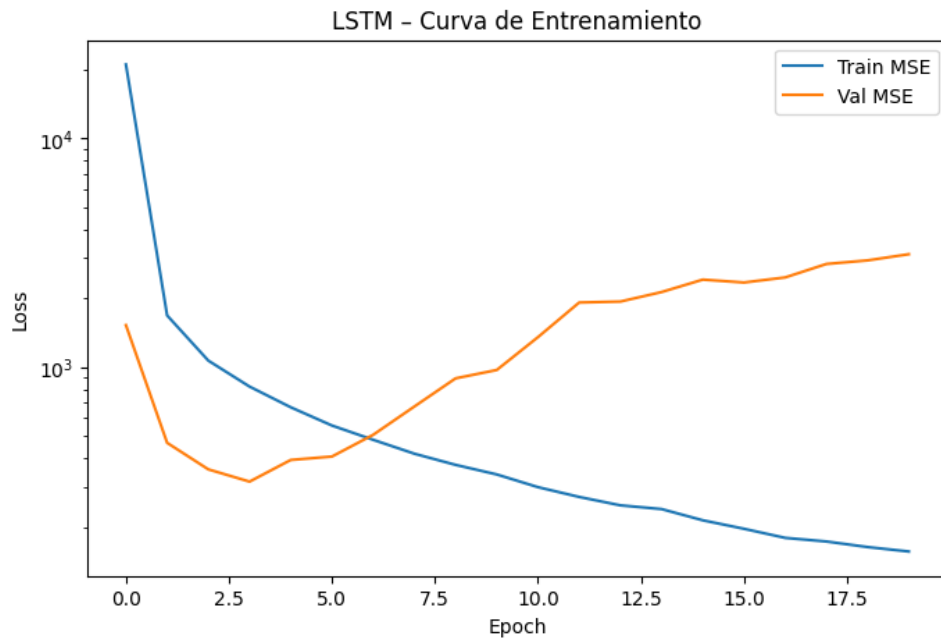


Figura 3: Curva de entrenamiento de la LSTM (MSE en escala logarítmica).

El análisis de estas tres figuras muestra que la combinación de profundidad moderada y estructura densa del MLP se adapta bien al tipo de relaciones presentes en los datos. La CNN 1D aporta una alternativa competitiva que también generaliza, aunque con errores levemente mayores. La LSTM, en cambio, parece no beneficiarse de la ventana temporal utilizada; el modelo intenta explotar dependencias que no se sostienen con claridad en el conjunto de validación y termina ajustándose en exceso al entrenamiento.

5. Relación entre valores reales y predichos

Además de las curvas de pérdida, se construyeron gráficos de dispersión entre los valores reales y los predichos por cada modelo sobre el conjunto de validación. En todos los casos se representan los puntos $(y_{\text{real}}, y_{\text{pred}})$ y se añade una línea diagonal que correspondería a una predicción perfecta.

Para el MLP, la Figura 4 muestra una nube de puntos extremadamente concentrada sobre la diagonal. La mayoría de las observaciones se sitúan prácticamente encima de la línea de referencia, con desviaciones pequeñas incluso en los valores más altos de NOx. Esta distribución refleja un ajuste muy preciso a lo largo de todo el rango dinámico del contaminante. No se observa sesgo sistemático: el modelo no tiende a sobreestimar ni subestimar regiones específicas de la escala de concentraciones.

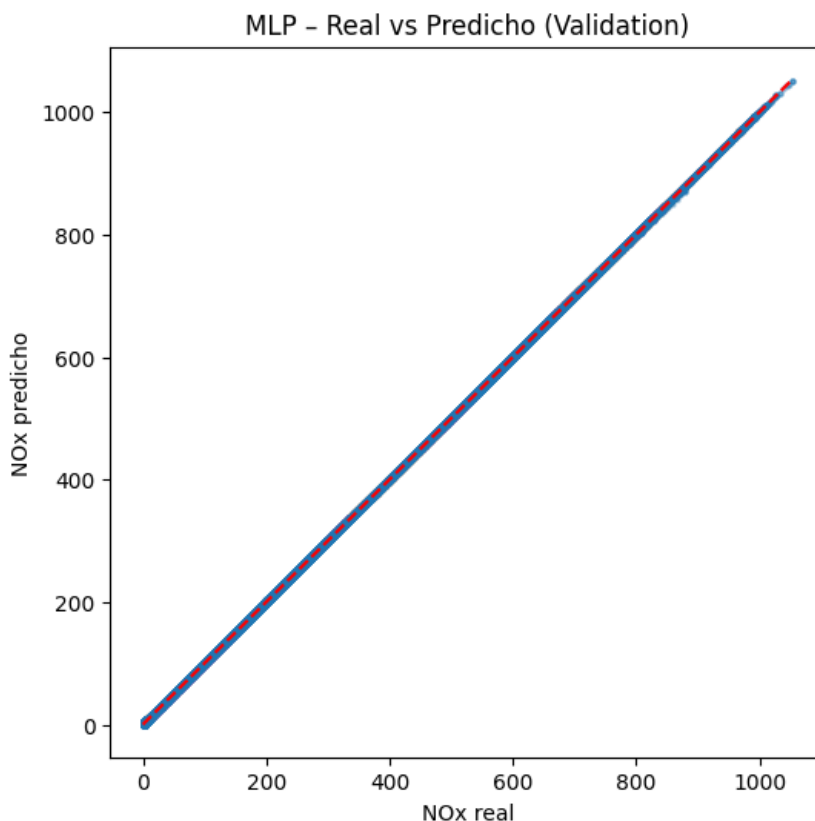


Figura 4: Relación entre NOx real y predicho por el MLP en validación.

En la CNN 1D, representada en la Figura 5, los puntos también se agrupan de manera notable alrededor de la diagonal. Existe algo más de dispersión que en el MLP, especialmente en los extremos del rango, pero la correlación visual sigue siendo muy alta. El patrón sugiere que la CNN captura bien la magnitud de los picos de NOx y reproduce de forma razonable las transiciones entre niveles bajos y altos. Algunas nubes de puntos se separan ligeramente por

encima o por debajo de la línea ideal, lo que coincide con el incremento de error observado en la curva de validación, pero sin llegar a degradar de forma marcada la capacidad predictiva global.

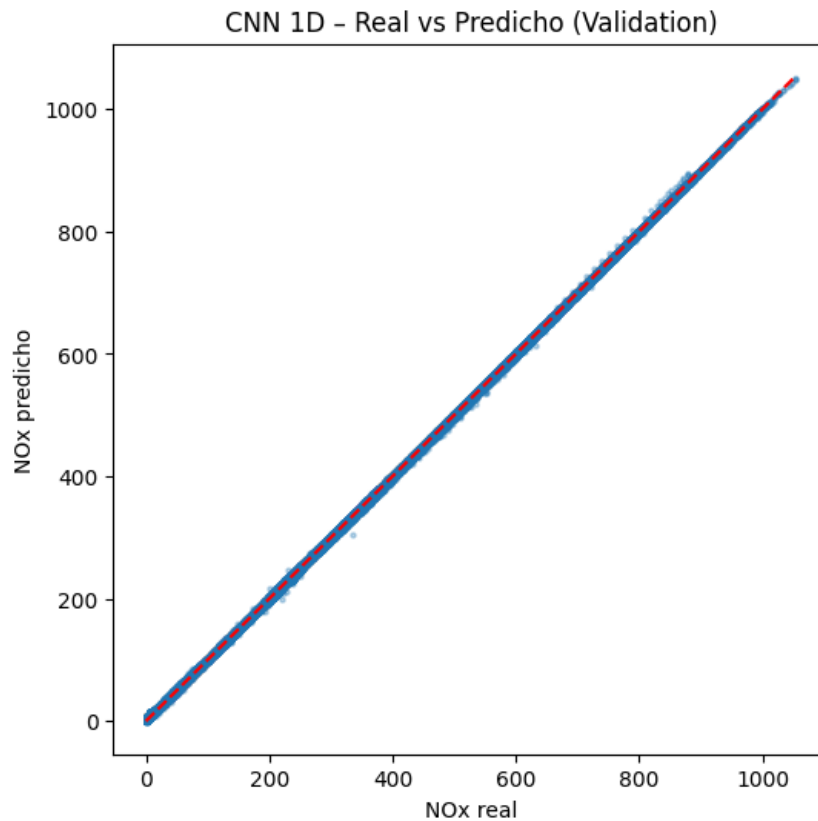


Figura 5: Relación entre NOx real y predicho por la CNN 1D en validación.

La Figura 6 ofrece una imagen distinta para la LSTM. En este caso, la nube de puntos tiende a situarse por debajo de la diagonal en las zonas de mayores concentraciones. El modelo reproduce con cierta fidelidad los valores bajos y medios de NOx, pero a medida que aumenta la concentración real, las predicciones se quedan rezagadas y no alcanzan los niveles más altos. Esto se traduce en una subestimación sistemática de los picos de emisión. El comportamiento encaja con la evolución de la pérdida de validación descrita antes: la red recurrente no sólo se ajusta en exceso al conjunto de entrenamiento, sino que además pierde capacidad para extrapolar adecuadamente en las regiones donde el contaminante presenta sus valores más extremos.

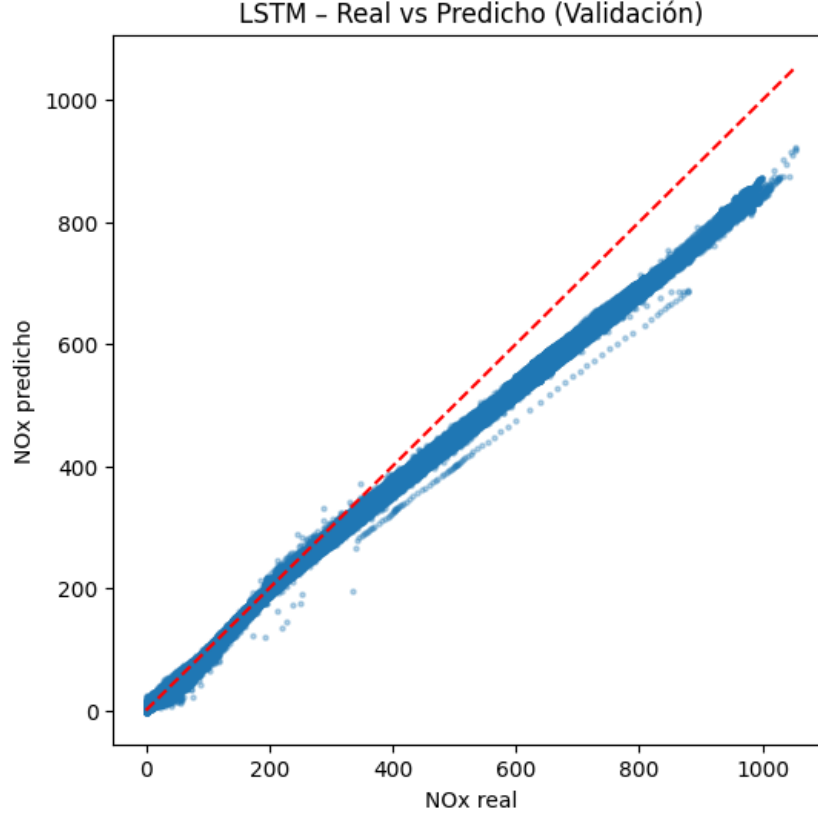


Figura 6: Relación entre NOx real y predicho por la LSTM en validación.

La comparación conjunta de estas tres figuras refuerza la idea de que, para este conjunto de datos, las arquitecturas densas y convolucionales funcionan mejor que la estructura recurrente. El MLP y la CNN representan de forma precisa tanto los valores de fondo como los picos, mientras que la LSTM tiende a comportarse de manera conservadora y recorta las emisiones más altas.

6. Resumen cuantitativo de las métricas

Los resultados numéricos obtenidos en el conjunto de prueba se recogen en la Tabla 1. Los valores de MAE y RMSE confirman lo observado en las figuras: el baseline que predice la media presenta errores muy altos, la regresión lineal mejora de forma notable estos valores, y los modelos MLP y CNN 1D reducen el error a órdenes de magnitud inferiores. La LSTM se sitúa en un punto intermedio: aunque el R^2 es elevado debido a la escala global de la variable, los errores absolutos son mayores que los de la CNN y el MLP.

Cuadro 1: Métricas de desempeño final en el conjunto de prueba.

Modelo	MAE	RMSE	R^2
Baseline (media)	272.13	335.20	-0.83
Regresión lineal	16.51	22.40	0.992
MLP	0.26	0.36	0.999998
CNN 1D	1.01	1.45	0.999893
LSTM	1.82	2.48	0.99989

El MLP se destaca como el modelo con mejor desempeño, al combinar una curva de entre-

namiento estable, una nube de puntos muy cercana a la diagonal real–predicho y errores medios reducidos. La CNN 1D aparece como una alternativa sólida, con ligeras penalizaciones en MAE y RMSE, pero con comportamiento visualmente consistente en las curvas y dispersores. La LSTM, aunque mantiene un R^2 alto debido a que captura la tendencia general de la variable, presenta un patrón de sobreajuste claro y una subestimación de los valores altos de NOx que limita su uso en escenarios donde los picos de emisión tienen relevancia.

7. Discusión general del trabajo

El proyecto permitió construir un pipeline completo de predicción de NOx que parte de archivos CSV crudos y culmina en un conjunto de modelos comparables entre sí. La decisión de fijar la variable objetivo en `NOx_wet_conc` asegura una interpretación directa de las salidas de los modelos, puesto que se trabaja con concentraciones húmedas medidas en condiciones cercanas al uso real del vehículo. La estrategia de normalización basada exclusivamente en estadísticas del conjunto de entrenamiento reduce el riesgo de optimismo artificial en las métricas.

La organización en notebooks numerados aportó claridad al proceso. Cada cuaderno se centró en una tarea específica: exploración, preprocesamiento, entrenamiento de un tipo de modelo o síntesis de resultados. Este esquema facilitó la depuración de errores y la evaluación de alternativas. Por ejemplo, la fase dedicada a la LSTM mostró de forma explícita cuánto se alejaban las curvas de entrenamiento y validación, y permitió descartar incrementos adicionales de complejidad en esa arquitectura.

Desde el punto de vista de modelos, el desempeño del MLP sugiere que la relación entre las variables de entrada y la concentración de NOx es altamente determinista y que puede capturarse con combinaciones no lineales relativamente sencillas. Las 46 características disponibles parecen contener suficiente información para reconstruir de manera precisa el nivel de emisiones sin necesidad de recurrir a mecanismos explícitos de memoria como los que ofrece una LSTM. La CNN 1D introduce la capacidad de analizar patrones locales dentro de una ventana temporal, lo cual también aporta una reconstrucción fiel, aunque ligeramente menos precisa que el MLP cuando se evalúa con las métricas elegidas.

La experiencia con la LSTM indica que no toda serie temporal multivariada se beneficia automáticamente de un tratamiento recurrente. En este caso, la ventana de 30 pasos quizás no sea lo bastante larga como para que las dependencias temporales justifiquen el costo de una capa recurrente, o bien las variables de entrada ya codifican de manera implícita el estado del sistema sin necesidad de integrar la historia completa. El aumento de la pérdida de validación y la subestimación sistemática de los picos visibles en la Figura 6 son señales claras de este comportamiento.

En términos de trabajo futuro, el pipeline actual podría ampliarse en varias direcciones. Una posibilidad sería investigar métodos de selección o extracción de características que reduzcan la dimensionalidad de entrada sin sacrificar capacidad predictiva, lo que podría favorecer arquitecturas más profundas o híbridas. Otra línea sería modificar la definición de las ventanas temporales para explorar horizontes más largos o enfoques de tipo *sequence-to-sequence*, donde el modelo prediga trayectorias completas de NOx en lugar de valores puntuales. También resultaría interesante evaluar la robustez de los modelos ante condiciones de operación diferentes de las usadas en el entrenamiento, por ejemplo nuevos trayectos o vehículos adicionales.

Referencias

- [1] Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. MIT Press, 2016.
- [2] Chollet, F. *Deep Learning with Python*. Manning Publications, 2017.