TerraWatt

Gestión inteligente del consumo eléctrico

Selección y Entrenamiento de Modelos para la Predicción del Precio de la Electricidad y del Consumo Energético

1. Predicción del Precio de la Electricidad con LSTM

Enfoque del Problema

La predicción del precio de la electricidad se considera un problema de series temporales, ya que los valores futuros dependen de la evolución histórica del sistema. Para este caso, se ha optado por un modelo de Redes Neuronales LSTM (Long Short-Term Memory), dado que tiene la capacidad de capturar relaciones temporales a largo plazo, un aspecto fundamental en este tipo de problemas.

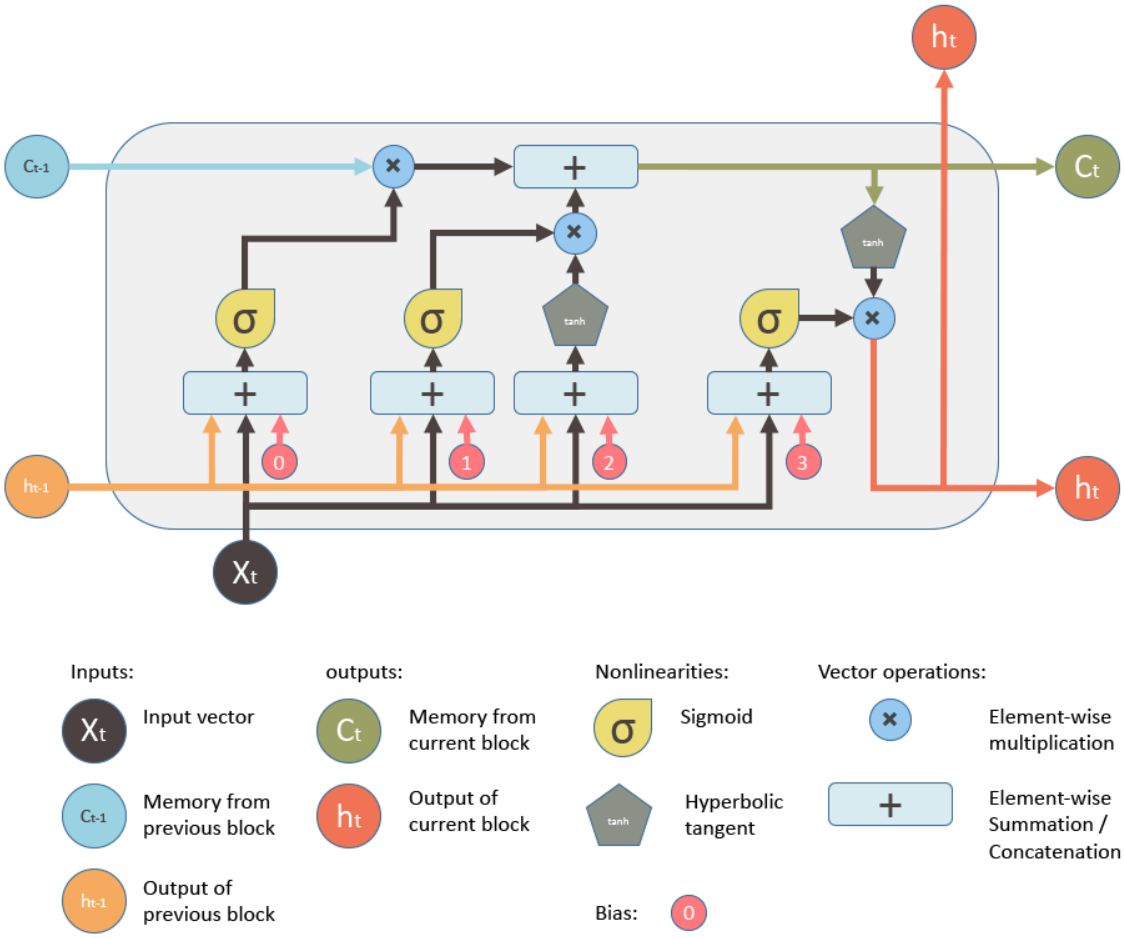
Justificación del Uso de LSTM

Las LSTM son una variante de las redes neuronales recurrentes diseñadas específicamente para manejar problemas con dependencia temporal extendida. En el contexto de la predicción del precio de la electricidad, estas redes son especialmente útiles porque permiten retener información pasada relevante sin los problemas de desaparición del gradiente que afectan a las RNN tradicionales. Además, pueden capturar dinámicas no lineales presentes en las fluctuaciones del precio y adaptarse a datos ruidosos y volátiles, lo cual es común en el mercado energético.

Arquitectura del Modelo

El modelo LSTM diseñado está compuesto por dos capas LSTM con 100 y 50 neuronas, respectivamente, utilizando la función de activación ReLU para mejorar la estabilidad del entrenamiento. La salida se obtiene a través de una capa densa con una sola neurona, la cual

predice el precio de la electricidad en €/MWh. En la primera capa LSTM se empleó la opción return_sequences = True para permitir la transferencia de memoria entre capas y mejorar la representación de secuencias largas.

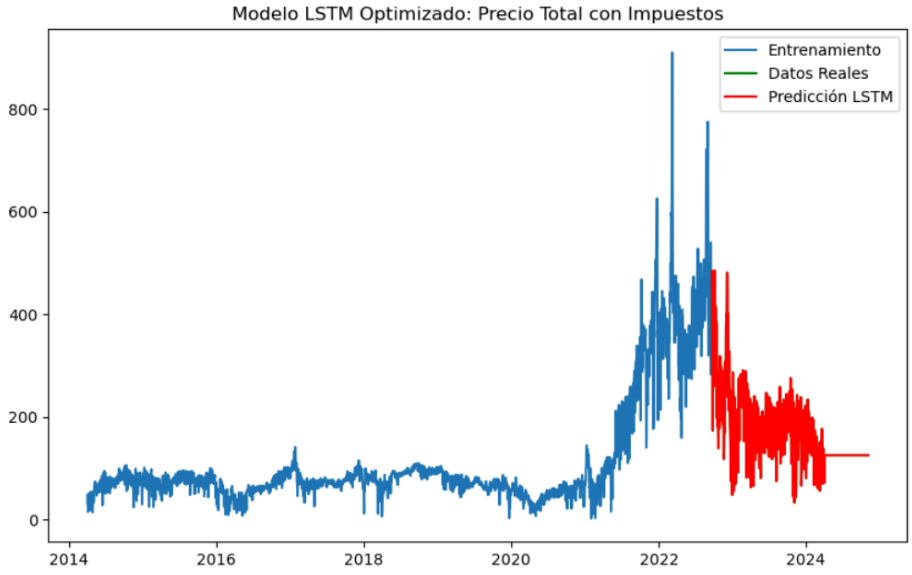


Proceso de Entrenamiento

El primer paso en el entrenamiento del modelo fue la carga y preprocesamiento de los datos. Para ello, se extrajeron los precios históricos de la electricidad desde un archivo CSV y se ordenaron cronológicamente. Luego, se dividió el conjunto de datos en un 80% para entrenamiento y un 20% para prueba, preservando el orden temporal. Posteriormente, se normalizaron los valores con el fin de mejorar la estabilidad del entrenamiento y se estructuraron los datos en el formato requerido por LSTM, con las dimensiones adecuadas de muestras, pasos de tiempo y características.

Durante la fase de entrenamiento, se utilizó el optimizador Adam, que proporciona una convergencia rápida y estable en modelos profundos. La función de pérdida seleccionada fue el error cuadrático medio (Mean Squared Error - MSE), dado que es la más utilizada en problemas de regresión. Para evitar el sobreajuste, se implementó la técnica de "Early

Stopping", que detiene el entrenamiento si el error de validación no mejora después de 10 épocas consecutivas. El modelo se entrenó durante 100 épocas con un tamaño de batch de 32, logrando un equilibrio entre precisión y tiempo de cómputo.



Evaluación y Resultados

Para medir el desempeño del modelo, se calcularon métricas como el error cuadrático medio raíz (Root Mean Squared Error - RMSE) y el coeficiente de determinación (R^2 Score). Además, se compararon las predicciones obtenidas con los valores reales para evaluar la capacidad del modelo de capturar patrones en la serie temporal. Los resultados mostraron que la red LSTM predijo con buena precisión las fluctuaciones del precio de la electricidad, logrando modelar tendencias, estacionalidad y dependencias a largo plazo en la serie de precios.

2. Predicción del Consumo Energético por Provincia

Factores que Afectan el Consumo Energético

El consumo energético depende de diversos factores, entre ellos las condiciones meteorológicas, las características de las viviendas y los factores demográficos. Entre las variables meteorológicas más influyentes se encuentran la temperatura, la radiación solar, la presión atmosférica y la velocidad del viento. En cuanto a las características de las viviendas, factores como el tipo de construcción, la superficie habitable y la potencia contratada pueden impactar significativamente en el consumo energético. Además, los factores demográficos, como el número de residentes en el hogar y sus hábitos de consumo, también juegan un papel importante.

Modelos Evaluados

Para abordar este problema, se analizaron distintos modelos de regresión con el objetivo de encontrar el más adecuado para cada provincia. El primer modelo considerado fue la regresión lineal, debido a su simplicidad e interpretabilidad, lo que permite identificar qué variables tienen mayor impacto en el consumo energético. Sin embargo, este modelo presenta limitaciones cuando los datos muestran relaciones no lineales.

Para superar esta limitación, se probó el modelo de Random Forest Regressor, basado en árboles de decisión, que tiene la capacidad de capturar relaciones complejas sin asumir una relación lineal entre las variables. Este modelo también es robusto ante valores atípicos y ruido en los datos, lo que lo hace especialmente útil en provincias con mayor variabilidad en el consumo.

Otro modelo evaluado fue el Gradient Boosting Regressor, el cual mejora la precisión respecto a Random Forest al construir árboles de decisión de forma secuencial, donde cada árbol corrige los errores del anterior. Este modelo es adecuado para conjuntos de datos más pequeños y permite captar patrones de consumo con mayor detalle. En provincias donde el consumo energético es más homogéneo, este enfoque obtuvo un mejor desempeño.

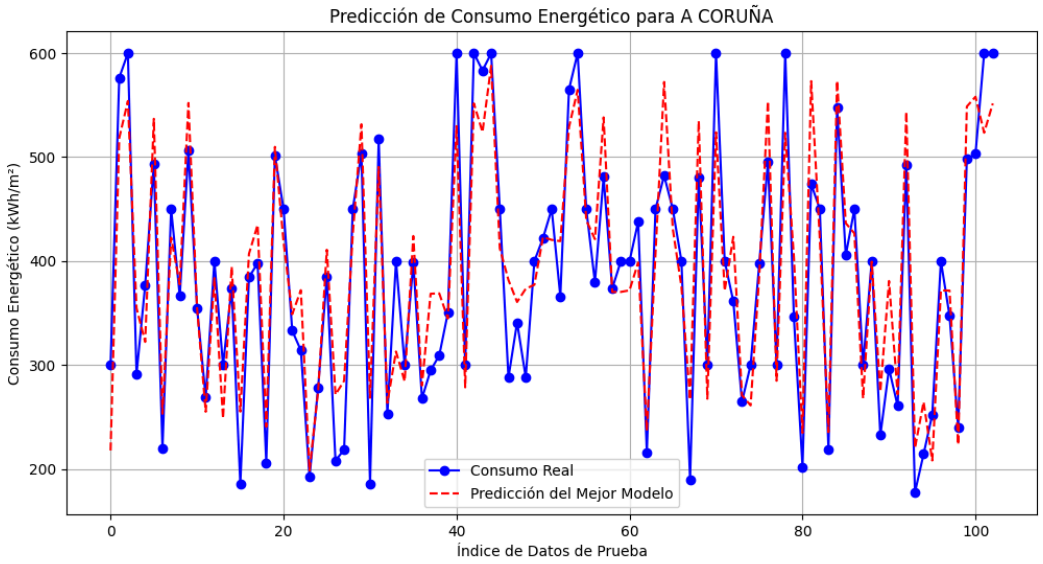
Finalmente, se consideró el modelo XGBoost (Extreme Gradient Boosting), una optimización del Gradient Boosting que se caracteriza por su eficiencia computacional, su manejo automático de valores faltantes y su capacidad de regularización, lo que ayuda a reducir el sobreajuste. XGBoost demostró ser el modelo más preciso en provincias con un mayor volumen de datos y alta variabilidad en el consumo energético.

Proceso de Entrenamiento

El preprocesamiento de datos incluyó la codificación de variables categóricas mediante One-Hot Encoding, en particular para variables como el tipo de vivienda. También se extrajeron características temporales, como el mes y el año, y se normalizaron las variables meteorológicas y de consumo utilizando StandardScaler.

Para entrenar y evaluar los modelos, los datos se dividieron en un 80% para entrenamiento y un 20% para prueba. Se entrenaron los cuatro modelos en cada provincia y se calcularon métricas de desempeño como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2). Finalmente, se seleccionó el modelo con el menor MSE en cada provincia para garantizar la mejor precisión posible.

	Model	MSE	R^2
1	Random Forest	1576.981749	0.897034
0	Linear Regression	1609.822065	0.894889
2	Gradient Boosting	1674.268976	0.890681
3	XGBoost	1887.254800	0.876775



Resultados y Selección del Mejor Modelo

Los resultados mostraron que no existe un modelo único que funcione mejor en todas las provincias, sino que la elección del modelo depende de las características de cada región. En algunas provincias, XGBoost obtuvo los mejores resultados gracias a su capacidad para modelar relaciones complejas en grandes volúmenes de datos. En otras, Random Forest mostró un mejor rendimiento, especialmente en conjuntos de datos más dispersos. Por otro lado, en provincias con datos más homogéneos, Gradient Boosting superó a los demás modelos.

Para garantizar un desempeño óptimo, se almacenó el modelo seleccionado de forma individual para cada provincia, asegurando así que cada conjunto de datos fuera tratado con el enfoque más adecuado.