

**Predicción del precio en bolsa de la energía eléctrica en Colombia mediante modelos estadísticos y redes neuronales**

Hoover Andrés Obando Toledo

Asesor

Edgard Augusto Trilleras Motta

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias de Ciencias Básicas, Tecnologías e Ingenierías ECBTI

Especialización en Ciencias de Datos y Analítica

2026

## Resumen

El precio en bolsa de la energía eléctrica en Colombia presenta alta volatilidad como consecuencia de la alta participación de generación hidroeléctrica en la matriz energética y la sensibilidad del sistema a variables climáticas y de demanda (Weron, 2014; XM S.A. E.S.P., 2023). Esta incertidumbre dificulta la planeación operativa, financiera y comercial de los agentes del mercado eléctrico. En respuesta a esta problemática, el presente trabajo tiene como propósito desarrollar y comparar modelos de predicción del precio en bolsa, utilizando enfoques estadísticos tradicionales como ARIMA ((Autoregressive Integrated Moving Average) y redes neuronales de tipo LSTM (Long Short-Term Memory), con el fin de determinar el modelo más preciso y robusto en el contexto colombiano.

Para ello, se recopilarán y analizarán datos históricos de precios, demanda, generación, reservas y variables hidrológicas provenientes del operador del sistema (XM S.A. E.S.P.), aplicando técnicas de análisis exploratorio, limpieza y transformación de datos. Posteriormente, se entrenarán y evaluarán los modelos utilizando métricas como el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Porcentual Absoluto Medio (MAPE), siguiendo lineamientos metodológicos ampliamente aceptados en la literatura (Hyndman & Athanasopoulos, 2018; Weron, 2014).

Se espera que este proyecto aporte tanto desde el punto de vista técnico como práctico, proporcionando un marco comparativo para la aplicación de modelos predictivos en mercados eléctricos. La propuesta se enmarca en la analítica energética y busca fortalecer las capacidades de planificación bajo escenarios de incertidumbre, en línea con las tendencias actuales de uso de inteligencia artificial en sistemas eléctricos.

***Palabras claves:*** Predicción, LSTM, ARIMA, mercado, XM, MAE, RMSE, MAPE

## Abstract

The spot price of electricity in Colombia exhibits high volatility due to the country's strong reliance on hydroelectric generation and the system's sensitivity to climatic and demand-side variables (Weron, 2014; XM S.A. E.S.P, 2023). This uncertainty poses challenges to the operational, financial, and commercial planning of market agents. In response, this study aims to develop and compare electricity price forecasting models using two distinct approaches: a traditional statistical model (ARIMA – Autoregressive Integrated Moving Average) and a deep learning-based model (LSTM – Long Short-Term Memory neural network), in order to identify the most accurate and robust model within the Colombian context.

Historical data from 2018 to 2023—covering spot prices, electricity demand, generation, hydrological contributions, and reservoir levels—will be collected from the national system operator (XM S.A. E.S.P.). Data will be preprocessed through exploratory analysis, cleaning, and transformation techniques. Both models will be trained and evaluated using widely accepted error metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), following established methodologies in the literature (Hyndman & Athanasopoulos, 2018; Weron, 2014).

This research is expected to provide both technical and practical insights by delivering a comparative framework for forecasting electricity market prices. The proposed work contributes to the field of energy analytics and aims to enhance planning capabilities under uncertainty, in alignment with current trends in the application of artificial intelligence to power systems.

**Keywords:** Forecasting, LSTM, ARIMA, electricity market, XM, MAE, RMSE, MAPE

## Tabla de Contenido

Introducción .....	9
Planteamiento del Problema .....	11
Justificación .....	13
Objetivos .....	15
Objetivo General .....	15
Objetivos Específicos.....	15
Marco Teórico.....	16
Modelos de la Familia ARIMA.....	16
Extensiones Estacionales y con Variables Exógenas: SARIMA y SARIMAX.....	17
Redes Neuronales LSTM .....	17
Marco Conceptual.....	20
Precio en Bolsa de la Energía Eléctrica .....	20
Series Temporales .....	20
Modelos de la Familia ARIMA.....	20
Redes Neuronales LSTM .....	21
Variables Exógenas.....	21
Métricas de Error.....	22
Estado del Arte y Revisión Bibliográfica .....	23
Modelos Estadísticos en la Predicción de Precios Eléctricos .....	24
Modelos Basados en Aprendizaje Automático y Profundo .....	25
Síntesis Comparativa y Aporte del Presente Trabajo.....	26
Metodología .....	27

Tipo de Investigación.....	27
Fuentes de Datos.....	27
Preprocesamiento de Datos.....	28
Modelos Implementados.....	28
Modelo SARIMAX.....	28
Red Neuronal LSTM.....	29
Evaluación de los Modelos.....	30
Herramientas y Software Utilizado.....	30
Análisis Exploratorio y Modelado de Series Temporales.....	32
Descripción de los Datos.....	32
Análisis Exploratorio de Datos (EDA).....	32
Valores Faltantes.....	33
Evolución del Precio Promedio Diario en Bolsa.....	34
Comportamiento de la Demanda Energética.....	35
Generación Eléctrica por Tecnología.....	35
Vertimientos Energéticos.....	36
Distribución del Precio Promedio Diario.....	37
Matriz de Correlación Entre Variables.....	37
Implementación de Modelos Predictivos.....	39
Modelo SARIMAX.....	39
Descomposición Estacional del Precio Promedio Diario.....	39
Serie Original.....	40
Tendencia (Trend).....	40

Estacionalidad (Seasonal).....	41
Residuos (Resid).....	41
Identificación y Configuración del Modelo SARIMAX.....	42
Ajuste y Validación del Modelo SARIMAX .....	42
Modelo LSTM Multivariado.....	42
Justificación del Enfoque .....	42
Preprocesamiento y Construcción de Secuencias .....	43
Arquitectura del Modelo .....	43
Entrenamiento y Evaluación .....	44
Resultados y Discusión.....	45
Criterio General de Evaluación y Comparabilidad Entre Modelos .....	45
Resultados del Modelo LSTM .....	46
Desempeño Cuantitativo .....	46
Justificación del Esquema One-Step-Ahead para el Modelo LSTM .....	48
Resultados del Modelo SARIMAX .....	48
Desempeño Cuantitativo .....	49
Justificación del Esquema de Backtesting por Bloques para el Modelo SARIMAX .....	51
Comparación Entre SARIMAX y LSTM .....	51
Coherencia Metodológica Entre Ambos Enfoques .....	52
Discusión de Resultados .....	52
Conclusiones.....	54
Recomendaciones y Trabajos Futuros .....	56
Referencias Bibliográficas .....	58

## Lista de Tablas

<b>Tabla 1</b> <i>Valores Faltantes por Variable</i> .....	33
<b>Tabla 2</b> <i>Comparación de Métricas Entre los Modelos Predictivos Implementados</i> .....	52

## Lista de Figuras

<b>Figura 1</b> <i>Tecnología de Generación Instalada en Colombia</i> .....	23
<b>Figura 2</b> <i>Evolución del Precio Promedio Diario en Bolsa</i> .....	34
<b>Figura 3</b> <i>Comportamiento de la Demanda Energética</i> .....	35
<b>Figura 4</b> <i>Generación Eléctrica por Tecnología</i> .....	36
<b>Figura 5</b> <i>Vertimientos Energéticos</i> .....	36
<b>Figura 6</b> <i>Distribución del Precio Promedio Diario</i> .....	37
<b>Figura 7</b> <i>Matriz de Correlación Entre Variables</i> .....	38
<b>Figura 8</b> <i>Descomposición Estacional del Precio Promedio Diario</i> .....	40
<b>Figura 9</b> <i>Predicción del Modelo LSTM (Normalizado) vs Precio Real en Bolsa</i> .....	47
<b>Figura 10</b> <i>Predicción del Modelo LSTM (Escala Real) vs Precio Real en Bolsa</i> .....	47
<b>Figura 11</b> <i>Predicción del Modelo SARIMAX (Normalizado) vs Precio Real en Bolsa</i> .....	50
<b>Figura 12</b> <i>Predicción del Modelo SARIMAX (Escala Real) vs Precio Real en Bolsa</i> .....	50

## Introducción

En el mercado eléctrico colombiano, el precio en bolsa constituye la principal señal económica que orienta la operación y la comercialización de la energía en el corto plazo. Este precio, calculado y administrado por el operador del sistema XM S.A. E.S.P., se determina de manera horaria y refleja el equilibrio entre la oferta de generación y la demanda de energía, sujeto a múltiples restricciones técnicas, operativas y económicas propias del sistema eléctrico (XM S.A. E.S.P., 2023).

No obstante, la alta dependencia del país de la generación hidroeléctrica —que representa más del 60 % de la capacidad instalada— expone al sistema a una elevada sensibilidad frente a condiciones climáticas adversas, particularmente fenómenos como El Niño y La Niña (UPME, 2022; Weron, 2014). Estas condiciones pueden generar episodios de alta volatilidad y cambios estructurales en el comportamiento del precio en bolsa, incrementando la incertidumbre para los agentes del mercado en sus decisiones comerciales, operativas y de planeación.

En este contexto, la predicción del precio en bolsa se convierte en una herramienta fundamental para la gestión del riesgo, la planeación estratégica y la formulación de políticas en el sector eléctrico. Tradicionalmente, este problema ha sido abordado mediante modelos estadísticos de series temporales, entre los cuales destacan los modelos autorregresivos integrados de media móvil (ARIMA) y sus extensiones estacionales, ampliamente estudiados bajo la metodología de Box y Jenkins (Box et al., 2016). Asimismo, se han empleado modelos aditivos como Prophet, que facilitan la captura de patrones estacionales y tendencias de largo plazo (Taylor & Letham, 2018). Sin embargo, estos enfoques presentan limitaciones para modelar relaciones no lineales y dinámicas complejas, especialmente en escenarios de alta volatilidad y cambios abruptos del mercado.

En años recientes, el avance del aprendizaje profundo ha impulsado el uso de modelos basados en redes neuronales recurrentes, en particular las redes Long Short-Term Memory (LSTM), las cuales han demostrado un desempeño superior en la predicción de series temporales complejas en el sector energético y en otros dominios con alta no linealidad (Hochreiter & Schmidhuber, 1997; Mishra et al., 2024). Diversos estudios reportan mejoras significativas en la precisión predictiva al comparar estos modelos con enfoques estadísticos tradicionales, especialmente cuando se integran variables exógenas relacionadas con la demanda, la generación y las condiciones hidrológicas del sistema (Li & Becker, 2021; Marcjasz et al., 2022).

En este marco, el presente trabajo tiene como objetivo evaluar y comparar el desempeño de un modelo estadístico perteneciente a la familia ARIMA, específicamente un modelo SARIMAX, frente a una red neuronal LSTM, para la predicción del precio promedio diario en bolsa del mercado eléctrico colombiano. La comparación se realiza considerando distintos esquemas de validación y métricas de error, con el propósito de identificar la estrategia de predicción más adecuada para un entorno caracterizado por alta volatilidad y complejidad operativa. Este trabajo se alinea con la creciente necesidad de incorporar técnicas de analítica avanzada e inteligencia artificial en los procesos de planeación y toma de decisiones del sector eléctrico.

## Planteamiento del Problema

El precio en bolsa de la energía eléctrica en Colombia es un indicador fundamental que orienta la operación diaria del Sistema Interconectado Nacional (SIN) y la toma de decisiones comerciales por parte de los agentes del mercado. Este precio, calculado de manera horaria por XM S.A. E.S.P., depende de múltiples factores, entre los que se incluyen la disponibilidad de generación, la demanda agregada, las restricciones operativas del sistema y, de forma particular, las condiciones hidrológicas del país. Dado que más del 60 % de la energía generada en Colombia proviene de fuentes hidroeléctricas, el sistema presenta una alta sensibilidad frente a fenómenos climáticos como El Niño, lo cual induce fuertes fluctuaciones en los precios de la energía y episodios de elevada volatilidad (XM S.A. E.S.P., 2023).

Esta volatilidad del precio en bolsa representa un desafío significativo para la gestión del riesgo, la programación de la generación y la negociación de contratos de mediano y largo plazo. Asimismo, dificulta la toma de decisiones informadas por parte de agentes generadores, comercializadores y grandes consumidores, quienes requieren herramientas confiables de pronóstico que les permitan anticipar el comportamiento del mercado y optimizar su operación. No obstante, la naturaleza dinámica y altamente no lineal del mercado eléctrico colombiano limita la efectividad de muchos enfoques tradicionales de predicción.

Históricamente, la predicción del precio de la energía se ha abordado mediante modelos estadísticos de series temporales, particularmente aquellos pertenecientes a la familia ARIMA, los cuales se basan en supuestos de linealidad y estacionariedad. Si bien estos modelos han demostrado utilidad en determinados contextos, sus supuestos estructurales no siempre se ajustan adecuadamente a las condiciones reales del mercado eléctrico colombiano, especialmente en

escenarios caracterizados por cambios estructurales, alta volatilidad y dependencia de factores climáticos.

En contraste, los modelos de aprendizaje automático y aprendizaje profundo, como las redes neuronales Long Short-Term Memory (LSTM), ofrecen la capacidad de capturar relaciones no lineales y patrones temporales complejos en series de datos extensas. Diversos estudios han reportado un mejor desempeño de estos modelos frente a enfoques estadísticos tradicionales en problemas de predicción de precios de energía (Hochreiter & Schmidhuber, 1997; Mishra et al., 2024). Sin embargo, en el contexto colombiano, los análisis que comparan de manera sistemática y bajo esquemas de validación coherentes el desempeño de modelos estadísticos avanzados (como SARIMAX, extensión de la familia ARIMA) frente a redes neuronales LSTM siguen siendo limitados.

En este sentido, surge la necesidad de evaluar y comparar distintos enfoques metodológicos para la predicción del precio en bolsa de la energía eléctrica en Colombia, considerando variables relevantes del mercado y aplicando criterios objetivos de validación. Una evaluación rigurosa de estos modelos permitiría identificar estrategias predictivas más adecuadas para apoyar la gestión operativa y comercial de los agentes del mercado, contribuyendo a una mayor eficiencia del sistema eléctrico y a una mejor planificación energética en el país.

## Justificación

El mercado eléctrico colombiano presenta particularidades estructurales que hacen crítica la predicción precisa del precio en bolsa, entre ellas su alta dependencia de la generación hidroeléctrica, la variabilidad climática y la sensibilidad de la demanda a factores económicos y sociales. Estos elementos introducen una elevada volatilidad en el precio horario y diario de la energía, lo cual representa un reto significativo para la planeación comercial, la programación de la generación y la formulación de estrategias de cobertura por parte de los agentes del mercado (XM S.A. E.S.P., 2023; Weron, 2014).

En este contexto, el desarrollo de modelos de predicción robustos y precisos resulta fundamental para reducir la incertidumbre asociada a los procesos de compra y venta de energía. Tradicionalmente, este problema ha sido abordado mediante modelos estadísticos de series temporales pertenecientes a la familia ARIMA, ampliamente reconocidos por su capacidad para capturar tendencias y estacionalidades en series univariadas (Box et al., 2016). Sin embargo, estos modelos presentan limitaciones al enfrentar comportamientos altamente no lineales y ante la necesidad de incorporar múltiples variables exógenas que influyen de manera significativa en la formación del precio en bolsa.

Frente a estas limitaciones, las redes neuronales recurrentes del tipo LSTM se han posicionado como una alternativa poderosa para la predicción de precios eléctricos, debido a su capacidad para modelar dependencias temporales de largo plazo y relaciones no lineales entre múltiples variables (Hochreiter & Schmidhuber, 1997; Mishra et al., 2024). Diversos estudios han reportado mejoras sustanciales en la precisión predictiva de estos modelos cuando se aplican a mercados eléctricos, especialmente al integrar variables como la demanda, la generación por

tecnología, los aportes hídricos y la disponibilidad del parque generador (Li & Becker, 2021; Lago et al., 2021).

El presente proyecto se justifica por su pertinencia técnica, su aplicabilidad práctica y su aporte académico. Desde el ámbito profesional, responde a una necesidad real observada en el entorno laboral del investigador, donde la toma de decisiones operativas y comerciales se encuentra fuertemente condicionada por la evolución del precio en bolsa. Desde el punto de vista técnico, la comparación entre un modelo estadístico avanzado de la familia ARIMA, específicamente un modelo SARIMAX, y una red neuronal LSTM permite establecer criterios objetivos de desempeño, precisión y viabilidad computacional en el contexto del mercado eléctrico colombiano.

Adicionalmente, este trabajo contribuye a la adopción de herramientas de analítica avanzada en los procesos de planeación energética nacional, alineándose con las tendencias globales de modernización del sector eléctrico y de toma de decisiones basada en datos. En este sentido, la investigación no solo busca comparar enfoques metodológicos, sino también aportar elementos que fortalezcan el análisis predictivo y la gestión informada en los distintos niveles del sector eléctrico colombiano (Marcjasz et al., 2022).

## Objetivos

### Objetivo General

Desarrollar y comparar modelos de predicción del precio en bolsa de la energía eléctrica en Colombia utilizando un modelo estadístico perteneciente a la familia ARIMA, específicamente un modelo SARIMAX, y una red neuronal de tipo LSTM, empleando variables históricas del mercado eléctrico entre los años 2000 y 2023, con el fin de determinar cuál enfoque presenta un mejor desempeño predictivo según las métricas de error MAE y RMSE.

### Objetivos Específicos

Recolectar y consolidar una base de datos multivariable correspondiente al período 2000–2023, que incluya el precio en bolsa de la energía eléctrica y variables relevantes del mercado eléctrico colombiano, tales como demanda, generación efectiva neta y vertimientos hídricos.

Implementar y ajustar un modelo estadístico de la familia ARIMA, específicamente un modelo SARIMAX, para la predicción del precio en bolsa de la energía eléctrica, considerando componentes autorregresivos, de media móvil y estacionales.

Desarrollar una red neuronal recurrente de tipo LSTM que permita la predicción del precio en bolsa a partir de múltiples variables exógenas del sistema eléctrico, tales como demanda, generación y vertimientos hídricos.

Evaluar y comparar el desempeño de los modelos desarrollados mediante métricas de error MAE y RMSE, así como a través del análisis gráfico del comportamiento temporal de las predicciones frente a los valores reales.

## Marco Teórico

El análisis y la predicción de series temporales constituyen una herramienta fundamental en el estudio de fenómenos energéticos, especialmente en mercados eléctricos, donde el precio en bolsa presenta una alta volatilidad asociada a factores como la demanda, la disponibilidad de generación y las condiciones hidrológicas del sistema (Weron, 2014). En este contexto, se han desarrollado diversos enfoques estadísticos y de aprendizaje automático para modelar y anticipar el comportamiento del precio de la energía.

### Modelos de la Familia ARIMA

Los modelos autorregresivos integrados de media móvil (ARIMA, por sus siglas en inglés AutoRegressive Integrated Moving Average) fueron desarrollados por Box y Jenkins y constituyen uno de los enfoques clásicos para el análisis de series temporales (Box et al., 2016). Estos modelos son de naturaleza lineal y univariada, y buscan capturar la dependencia temporal de una serie estacionaria o transformada en estacionaria mediante diferenciación. Un modelo ARIMA se compone de tres elementos principales:

- AR(p): parte autorregresiva, donde el valor actual depende de sus valores pasados.
- I(d): número de diferenciaciones necesarias para hacer la serie estacionaria.
- MA(q): parte de media móvil, que modela la dependencia con errores pasados

Una serie  $y_t$  modelada por un ARIMA(p,d,q) se expresa como:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t$$

Donde:

- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ : operador autorregresivo
- $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ : operador de media móvil
- $B$ : operador de rezago,  $By_t = y_{t-1}$

- $\varepsilon_t$ : ruido blanco

Estos modelos han sido aplicados con éxito en la predicción de precios eléctricos en distintos mercados (Zhao et al., 2018; Muñoz-Santiago et al., 2017). No obstante, su principal limitación radica en la imposibilidad de capturar relaciones no lineales complejas y de incorporar de manera directa múltiples variables explicativas del sistema.

### ***Extensiones Estacionales y con Variables Exógenas: SARIMA y SARIMAX***

Con el fin de superar algunas de las limitaciones del modelo ARIMA, se han propuesto extensiones que permiten capturar estacionalidades y efectos multivariados. El modelo SARIMA (Seasonal ARIMA) incorpora componentes autorregresivos y de media móvil estacionales, lo que resulta especialmente útil para series con patrones periódicos, como las series de precios de energía.

Adicionalmente, el modelo SARIMAX (Seasonal ARIMA with exogenous variables) permite incluir variables exógenas que influyen en la variable objetivo, tales como la demanda, la generación por tecnología o los aportes hídricos. Esta extensión amplía la capacidad explicativa del modelo y lo hace más adecuado para sistemas complejos como el mercado eléctrico colombiano, donde el precio depende de múltiples factores operativos y climáticos.

A pesar de estas mejoras, los modelos SARIMA y SARIMAX conservan una estructura esencialmente lineal, lo que puede limitar su desempeño en escenarios caracterizados por alta volatilidad, cambios estructurales y relaciones no lineales entre variables.

### **Redes Neuronales LSTM**

Las redes neuronales recurrentes (RNN) constituyen una clase de modelos diseñados para el procesamiento de datos secuenciales, al incorporar mecanismos de retroalimentación que permiten modelar dependencias temporales. Sin embargo, las RNN tradicionales presentan

dificultades para capturar dependencias de largo plazo debido a problemas de desvanecimiento o explosión del gradiente durante el entrenamiento (Hochreiter & Schmidhuber, 1997).

Para abordar estas limitaciones, se desarrollaron las redes Long Short-Term Memory (LSTM), las cuales incorporan una estructura de memoria controlada mediante puertas que regulan el flujo de información. Una celda LSTM está compuesta por tres mecanismos principales: la puerta de olvido, la puerta de entrada y la puerta de salida, los cuales permiten decidir qué información se descarta, se actualiza o se transmite al siguiente estado.

Las ecuaciones internas de una celda LSTM en el instante  $t$  pueden expresarse como:

- Puerta de olvido:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- Puerta de entrada:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

- Estado de la celda:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

- Puerta de salida:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

- Estado oculto:

$$h_t = o_t \cdot \tanh(C_t)$$

Donde:

- $x_t$ : entrada en el tiempo  $t$
- $h_{t-1}$ : salida del paso anterior.
- $C_t$ : memoria interna.
- $W, b$ : pesos y sesgos entrenables.

- $\sigma$ : función sigmoide.
- $\tanh$ : tangente hiperbólica.

Las redes LSTM han demostrado un alto desempeño en la predicción de precios eléctricos, especialmente cuando se emplean como modelos multivariantes que integran información exógena del sistema, tales como demanda, generación, aportes hídricos y condiciones operativas (Mishra et al., 2024; Lago et al., 2021). Su capacidad para modelar relaciones no lineales y dependencias temporales de largo plazo las convierte en una alternativa robusta frente a los modelos estadísticos tradicionales.

## **Marco Conceptual**

A continuación, se presentan los principales conceptos que sustentan el desarrollo de este proyecto de predicción del precio en bolsa de la energía eléctrica en Colombia:

### **Precio en Bolsa de la Energía Eléctrica**

Es el valor económico que se determina para la energía transada en el Mercado Mayorista de Energía en Colombia. Este precio se calcula a partir de la interacción entre la oferta presentada por los generadores y la demanda de los comercializadores y usuarios no regulados. El Operador del Sistema Interconectado Nacional, XM S.A. E.S.P., publica el precio horario en bolsa, el cual constituye un referente esencial para la liquidación de transacciones, el cubrimiento de contratos y la planeación financiera de los agentes del mercado (XM S.A. E.S.P., 2023).

### **Series Temporales**

Son secuencias de observaciones de una variable ordenadas cronológicamente. El análisis de series temporales permite identificar y modelar patrones de comportamiento que pueden incluir tendencias de largo plazo, estacionalidades periódicas, ciclos y componentes aleatorios. La predicción de series temporales tiene una importancia estratégica en diversos sectores económicos, especialmente en el sector energético, debido a su impacto directo en la toma de decisiones operativas, comerciales y regulatorias (Hyndman & Athanasopoulos, 2018).

### **Modelos de la Familia ARIMA**

Los modelos autorregresivos integrados de media móvil (ARIMA) fueron desarrollados por Box y Jenkins y constituyen una de las técnicas más utilizadas para la predicción de series temporales (Box et al., 2016). Estos modelos combinan tres componentes fundamentales: la parte autorregresiva (AR), que modela la relación con valores pasados de la serie; la integración (I),

que permite diferenciar la serie para alcanzar la estacionariedad; y la media móvil (MA), que incorpora la dependencia con errores pasados.

Dentro de esta familia se incluyen extensiones como los modelos SARIMA y SARIMAX, que permiten capturar patrones estacionales y la influencia de variables exógenas, respectivamente. Estos modelos han sido ampliamente utilizados en aplicaciones energéticas debido a su simplicidad, interpretabilidad y bajo costo computacional, aunque presentan limitaciones para modelar relaciones no lineales complejas.

### **Redes Neuronales LSTM**

Las redes neuronales de memoria a largo y corto plazo (Long Short-Term Memory, LSTM) fueron propuestas por Hochreiter y Schmidhuber (1997) como una extensión de las redes neuronales recurrentes. Estas redes incorporan mecanismos internos de control, conocidos como puertas de entrada, olvido y salida, que permiten conservar y actualizar información relevante a lo largo de secuencias temporales extensas.

Gracias a esta arquitectura, las redes LSTM son capaces de modelar relaciones no lineales y dependencias temporales de largo plazo que resultan difíciles de capturar mediante modelos estadísticos convencionales. En el sector energético, las LSTM han sido aplicadas exitosamente en la predicción de precios, demanda y generación, mostrando mejoras significativas en precisión cuando se integran múltiples variables explicativas (Mishra et al., 2024; Li & Becker, 2021).

### **Variables Exógenas**

Las variables exógenas corresponden a factores externos que no dependen directamente de la variable objetivo, pero que influyen de manera significativa en su comportamiento. En el contexto del precio en bolsa de la energía eléctrica, estas variables incluyen la demanda de

energía, la generación total y por tecnología (hidroeléctrica, térmica), los aportes hídricos a los embalses y variables climáticas como la precipitación o la temperatura. La incorporación de variables exógenas permite mejorar la capacidad explicativa y predictiva de los modelos, especialmente en mercados eléctricos caracterizados por alta complejidad y volatilidad (Marcjasz et al., 2022).

### **Métricas de Error**

Las métricas de error son indicadores cuantitativos utilizados para evaluar el desempeño de los modelos de predicción. En este trabajo se emplean principalmente las siguientes métricas:

**Error Absoluto Medio (MAE):** mide el promedio de las diferencias absolutas entre los valores predichos y los valores observados.

**Raíz del Error Cuadrático Medio (RMSE):** calcula la raíz cuadrada del promedio de los errores al cuadrado, penalizando en mayor medida los errores de gran magnitud.

Estas métricas permiten comparar objetivamente el desempeño de los modelos y determinar cuál ofrece predicciones más precisas y confiables en el contexto analizado (Hyndman & Athanasopoulos, 2018).




## Estado del Arte y Revisión Bibliográfica

El mercado eléctrico colombiano opera bajo un esquema de liberalización y libre competencia establecido a partir de las Leyes 142 y 143 de 1994, mediante las cuales se creó el Mercado de Energía Mayorista (MEM). Este mercado permite la libre participación de los agentes en las actividades de generación y comercialización, mientras que la transmisión y distribución permanecen reguladas. En este contexto, coexisten dos tipos de mercados: el mercado regulado, basado en contratos bilaterales de largo plazo, y el mercado no regulado o bolsa de energía, donde la asignación se realiza mediante un mecanismo de subasta diaria entre los generadores (Barrientos et al., 2012).

En el mercado de corto plazo, el precio de la energía refleja tanto factores técnicos — como el tipo de tecnología de generación, la disponibilidad de las plantas y la localización nodal— como factores de mercado asociados a la oferta y la demanda. La estructura de generación del Sistema Interconectado Nacional (SIN) se caracteriza por una alta participación de la generación hidráulica, la cual representa la mayor proporción de la capacidad instalada del país (XM, 2024). En consecuencia, el precio en bolsa está fuertemente influenciado por las condiciones hidrológicas, lo que introduce altos niveles de volatilidad y complejidad en su comportamiento.

### Figura 1

#### *Tecnología de Generación Instalada en Colombia*

Tecnología	Capacidad Efectiva Neta (MW)	%
 Hidráulica	13.220	63,7
 Térmica	6.274	30,2
 Solar	1.273	6,1

*Nota.* Tomada de XM 2024.

## Modelos Estadísticos en la Predicción de Precios Eléctricos

Desde el ámbito econométrico y financiero, los modelos estadísticos han sido ampliamente utilizados para la predicción de variables económicas y energéticas. En particular, los modelos de la familia ARIMA han demostrado ser herramientas eficaces para capturar patrones lineales, tendencias y estacionalidades en series temporales. No obstante, estos modelos suelen presentar limitaciones cuando se enfrentan a relaciones no lineales o a la incorporación de múltiples variables explicativas, fenómeno conocido como la “maldición de la dimensionalidad”.

En el contexto colombiano, Gómez-Cano et al. (2020) desarrollaron modelos estadísticos para la predicción del precio en bolsa, analizando la influencia de variables macroeconómicas y técnicas como el PIB, el precio del gas, la disponibilidad de generación, la demanda máxima y la capacidad efectiva neta del sistema. Los autores encontraron correlaciones positivas con variables de demanda y generación, y correlaciones negativas con variables hidrológicas como el nivel de los embalses. En este estudio se emplearon modelos SARIMA y SARIMAX bajo la metodología de Box-Jenkins.

De manera similar, González Pérez y Urbano Buriticá (2022) evaluaron modelos de corto plazo para la proyección del precio en bolsa colombiano, integrando variables como demanda, vertimientos, reservas, generación por tecnología y aportes hídricos. Los autores compararon enfoques ARIMA, Prophet, SARIMAX y redes neuronales artificiales, evidenciando que los modelos estadísticos presentan ventajas en términos de interpretabilidad y menor costo computacional, mientras que los modelos basados en redes neuronales ofrecen mayor flexibilidad para capturar comportamientos complejos.

A nivel internacional, Zhao et al. (2018) demostraron que la incorporación de variables exógenas en modelos ARIMA puede mejorar el desempeño predictivo, aunque resaltaron que

estos modelos tienden a perder precisión ante eventos abruptos, como shocks regulatorios o condiciones climáticas extremas.

### **Modelos Basados en Aprendizaje Automático y Profundo**

Con el avance del aprendizaje automático, los modelos basados en redes neuronales han ganado protagonismo en la predicción de precios eléctricos. En particular, las redes neuronales con memoria a largo y corto plazo (LSTM) han mostrado un desempeño superior en la captura de relaciones no lineales y dependencias temporales complejas.

Geller (2021) realizó una comparación entre modelos estadísticos tradicionales y redes LSTM en la predicción de índices financieros, evidenciando mejoras significativas en precisión al emplear modelos neuronales. De manera similar, da Matta et al. (2021) evaluaron el desempeño de redes neuronales frente a modelos de regresión en conjuntos de datos macroeconómicos reales de Estados Unidos, concluyendo que las redes neuronales presentan una mayor capacidad de adaptación a cambios estructurales.

En el ámbito energético, Mishra et al. (2024) compararon modelos clásicos y redes LSTM para la predicción de precios eléctricos en el Reino Unido, concluyendo que las redes neuronales superan a los modelos estadísticos, especialmente en horizontes de corto plazo. Li y Becker (2021) integraron modelos LSTM con selección de características para predecir precios diarios en mercados acoplados, obteniendo errores significativamente menores. Asimismo, Marcjasz et al. (2022) propusieron modelos neuronales probabilísticos capaces de estimar distribuciones completas del precio, ampliando el alcance de la predicción más allá del valor esperado.

## **Síntesis Comparativa y Aporte del Presente Trabajo**

La revisión bibliográfica permite concluir que los modelos estadísticos de la familia ARIMA y SARIMAX continúan siendo herramientas relevantes para la predicción de precios eléctricos, especialmente por su interpretabilidad y eficiencia computacional. Sin embargo, su desempeño se ve limitado frente a mercados altamente volátiles y no lineales. Por otro lado, las redes LSTM destacan por su capacidad para aprender patrones complejos y adaptarse a cambios abruptos, aunque requieren mayores volúmenes de datos y recursos computacionales.

En el contexto colombiano, aún son escasos los estudios que realicen comparaciones sistemáticas y metodológicamente consistentes entre modelos estadísticos tradicionales y modelos de aprendizaje profundo utilizando series históricas largas y variables multivariantes públicas. En este sentido, el presente proyecto busca contribuir a cerrar esta brecha mediante una comparación robusta entre un modelo SARIMAX y una red neuronal LSTM, utilizando datos reales del mercado eléctrico colombiano entre los años 2000 y 2023 e incorporando variables relevantes como demanda, generación, vertimientos y aportes hídricos.

## **Metodología**

La metodología de este proyecto se enmarca en un enfoque cuantitativo, comparativo y aplicado, orientado a evaluar el desempeño de modelos estadísticos tradicionales y modelos de aprendizaje profundo para la predicción del precio en bolsa de la energía eléctrica en Colombia. El estudio se desarrolló en varias fases que incluyen la recopilación de datos, el preprocesamiento, la implementación de modelos predictivos, la evaluación de resultados y la comparación entre enfoques.

### **Tipo de Investigación**

La investigación es de tipo exploratoria y aplicada. Es exploratoria en la medida en que busca analizar y contrastar el comportamiento de modelos estadísticos clásicos frente a modelos basados en redes neuronales profundas en el contexto específico del mercado eléctrico colombiano, caracterizado por alta volatilidad y no linealidad. Asimismo, es aplicada, ya que los modelos desarrollados están orientados a resolver un problema real de predicción del precio en bolsa, con utilidad directa para la toma de decisiones operativas y comerciales en el sector energético.

### **Fuentes de Datos**

Los datos utilizados en este estudio fueron obtenidos de fuentes oficiales del sector eléctrico colombiano:

XM S.A. E.S.P.: Se recolectaron series históricas del precio en bolsa de la energía eléctrica, demanda agregada, generación efectiva neta por tecnología y variables hidrológicas asociadas al sistema, con frecuencia diaria, para el período comprendido entre los años 2000 y 2023.

Estas fuentes garantizan consistencia, trazabilidad y representatividad de la información utilizada en el modelado.

### **Preprocesamiento de Datos**

Antes de la implementación de los modelos predictivos, se realizó un proceso de análisis exploratorio y preprocesamiento de los datos, con el fin de garantizar su calidad y adecuación metodológica. Este proceso incluyó:

**Limpieza de datos:** Identificación y tratamiento de valores faltantes, atípicos o inconsistentes. En caso necesario, se aplicaron técnicas de imputación temporal para preservar la continuidad de las series.

**Transformación de las series temporales:** Para los modelos estadísticos, se verificó la estacionariedad de la serie objetivo mediante pruebas estadísticas y se aplicaron transformaciones de diferenciación cuando fue requerido.

**Normalización de variables:** Para el modelo LSTM, las variables de entrada fueron normalizadas en el rango  $[0, 1]$  utilizando el método Min–Max, con el fin de facilitar el proceso de entrenamiento y mejorar la estabilidad numérica del modelo. La normalización se realizó evitando el uso de información futura (data leakage).

### **Modelos Implementados**

#### ***Modelo SARIMAX***

Si bien inicialmente se consideró el uso de modelos ARIMA univariados, durante el desarrollo del proyecto se optó por implementar un modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables), dado que este permite incorporar explícitamente componentes estacionales y variables exógenas relevantes para el mercado eléctrico colombiano.

El procedimiento seguido para el modelo SARIMAX fue el siguiente:

Selección de parámetros: Los órdenes del modelo  $(p,d,q)$  y  $(P,D,Q,s)$  fueron definidos con base en criterios de información como AIC y BIC, así como en análisis de autocorrelación y autocorrelación parcial.

Esquema de evaluación: El modelo fue evaluado mediante un esquema de backtesting con reentrenamiento periódico por bloques de 30 días, en el cual el modelo se recalibra utilizando toda la información histórica disponible hasta el inicio de cada bloque y genera predicciones para el período siguiente.

Justificación computacional: Aunque conceptualmente es posible implementar un esquema one-step-ahead diario, este enfoque resulta computacionalmente costoso para modelos SARIMAX debido a la reestimación frecuente de parámetros mediante máxima verosimilitud. Por esta razón, se adoptó el esquema por bloques, que reduce significativamente el tiempo de ejecución y representa un escenario operativo realista para modelos estadísticos.

### ***Red Neuronal LSTM***

La red neuronal LSTM fue implementada con el objetivo de capturar relaciones no lineales y dependencias de largo plazo en el comportamiento del precio en bolsa, incorporando múltiples variables exógenas. Las principales características de la implementación fueron:

Arquitectura del modelo: Red LSTM multivariada con dos capas ocultas de unidades LSTM, seguidas de capas de regularización (Dropout) y una capa densa de salida.

Ventana temporal: Se utilizó una ventana deslizante de 30 días para la construcción de las secuencias de entrada.

Entrenamiento:

- Optimizador: Adam.

- Función de pérdida: Mean Squared Error (MSE).
- Tamaño de lote: 32 observaciones.
- Estrategia de regularización: Early stopping para evitar sobreajuste.

Esquema de evaluación: Se utilizó un enfoque one-step-ahead, en el cual el modelo es entrenado una única vez y genera predicciones diarias incorporando el valor real observado en cada paso (teacher forcing). Este esquema resulta computacionalmente eficiente para redes LSTM y permite obtener métricas estables y comparables.

### **Evaluación de los Modelos**

El desempeño de los modelos SARIMAX y LSTM fue evaluado mediante métricas de error ampliamente utilizadas en la literatura:

- MAE (Mean Absolute Error).
- RMSE (Root Mean Squared Error).

Las métricas fueron calculadas tanto en escala normalizada como en escala real (\$/kWh), con el fin de garantizar comparabilidad entre modelos y facilitar la interpretación práctica de los resultados. Adicionalmente, se realizaron gráficos de comparación entre valores reales y predichos para analizar visualmente el comportamiento temporal y la capacidad de ajuste de cada modelo.

Cabe resaltar que, aunque inicialmente se consideró el uso de la métrica MAPE, esta no fue incluida en la evaluación final debido a la presencia de valores cercanos a cero y alta volatilidad, lo que puede generar distorsiones en su interpretación.

### **Herramientas y Software Utilizado**

- Lenguaje de programación: Python 3.8+.

- Librerías principales: Pandas, NumPy, statsmodels (SARIMAX), TensorFlow/Keras (LSTM), scikit-learn (preprocesamiento y métricas).
- Entorno de desarrollo: Jupyter Notebook

## **Análisis Exploratorio y Modelado de Series Temporales**

Antes de aplicar los modelos de predicción, se llevó a cabo un análisis exploratorio exhaustivo con el objetivo de comprender la estructura y el comportamiento de las variables involucradas, identificar patrones relevantes y preparar adecuadamente los datos para su posterior modelamiento.

### **Descripción de los Datos**

El presente estudio se basa en una serie temporal diaria del precio en bolsa de la energía eléctrica en Colombia, comprendida entre los años 2000 y 2023. Esta información fue recolectada de fuentes oficiales del mercado eléctrico colombiano, incluyendo registros históricos publicados por XM S.A. E.S.P., el operador del Sistema Interconectado Nacional y administrador del Mercado de Energía Mayorista (MEM).

La variable objetivo del estudio es el “Precio Promedio Diario de Bolsa” expresado en pesos colombianos por kilovatio-hora (\$/kWh). Adicionalmente, se dispuso de variables exógenas relevantes como: demanda diaria del Sistema Interconectado Nacional (SIN), generación por fuente (hidráulica, térmica, eólica, solar), energía no atendida, vertimientos de energía, entre otras, permitiendo un enfoque multivariado en el modelado predictivo.

### **Análisis Exploratorio de Datos (EDA)**

Se realizó un análisis exploratorio exhaustivo para comprender la dinámica y distribución de las variables. Inicialmente, se identificaron valores faltantes en múltiples columnas, los cuales fueron imputados mediante interpolación lineal y técnicas de forward/backward fill, garantizando la coherencia temporal de la serie.

A través de visualizaciones como histogramas, gráficos de líneas, distribuciones kernel y matrices de correlación, se evidenciaron los siguientes patrones:

- Tendencia creciente en la demanda energética.
- Patrones estacionales en la generación hidráulica, típicamente relacionados con los ciclos hidrológicos.
- Eventos extremos en el precio de bolsa, asociados a periodos de alta escasez hídrica (Fenómeno de El Niño).
- Correlaciones positivas entre el precio y la generación térmica, y negativas con los vertimientos hidráulicos.

Este análisis permitió entender el comportamiento conjunto de las variables y su potencial relación causal con el precio energético. A continuación, se presenta un análisis más detallado del EDA realizado.

### ***Valores Faltantes***

El conjunto de datos presenta algunas ausencias puntuales, especialmente en variables de generación por tecnología (eólica, térmica, hidráulica) y vertimientos energéticos. Se detectaron, por ejemplo, 376 registros faltantes en la generación eólica e hidráulica, así como más de 3000 ausencias en los vertimientos de energía y volumen. Estos vacíos fueron imputados utilizando el método de forward fill (relleno hacia adelante), garantizando la continuidad temporal de las series (Wickham & Grolemund, 2017).

### **Tabla 1**

#### *Valores Faltantes por Variable*

Variable	Número de datos faltantes
Fecha	11
Promedio	9
Demanda Energia SIN kWh	12
Demanda No Atendida kWh	499

Variable	Número de datos faltantes
COGENERADOR	376
EOLICA	376
HIDRAULICA	376
TERMICA	376
Vertimientos Volumen miles m3	3034
Vertimientos Energía kWh	3034
SOLAR_TOTAL	0

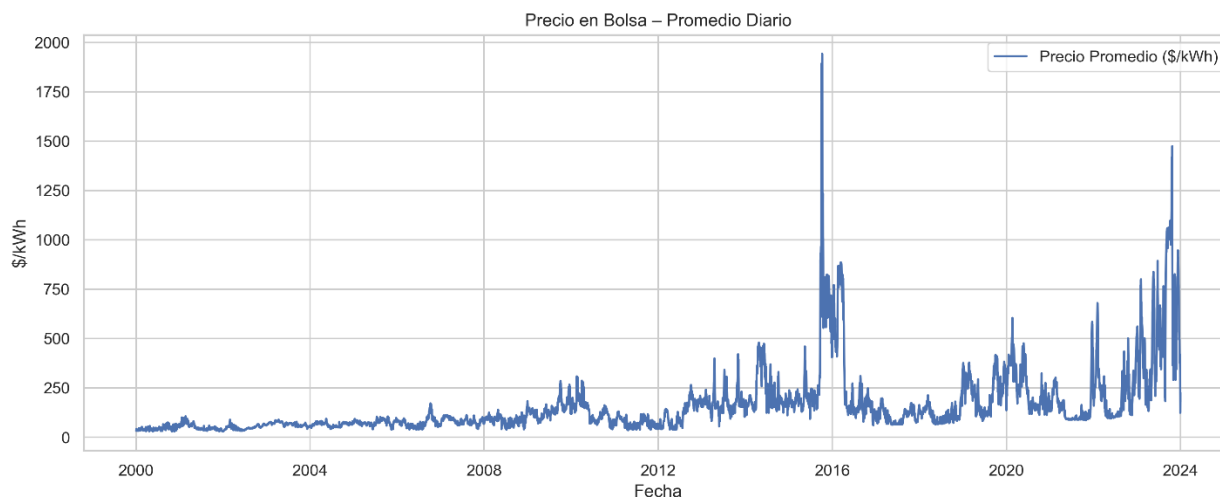
*Nota.* La tabla contiene la cantidad de valores faltantes por variable utilizada para el análisis

### ***Evolución del Precio Promedio Diario en Bolsa***

En la Figura 2 se visualiza la serie histórica del precio promedio diario en bolsa desde el año 2000 hasta inicios de 2024. Se evidencia un comportamiento altamente volátil, con incrementos abruptos durante eventos críticos del sistema (por ejemplo, el fenómeno de El Niño en 2015-2016) y una clara tendencia alcista a partir del año 2022, lo que podría atribuirse al aumento de los costos marginales y cambios regulatorios.

### **Figura 2**

#### *Evolución del Precio Promedio Diario en Bolsa*

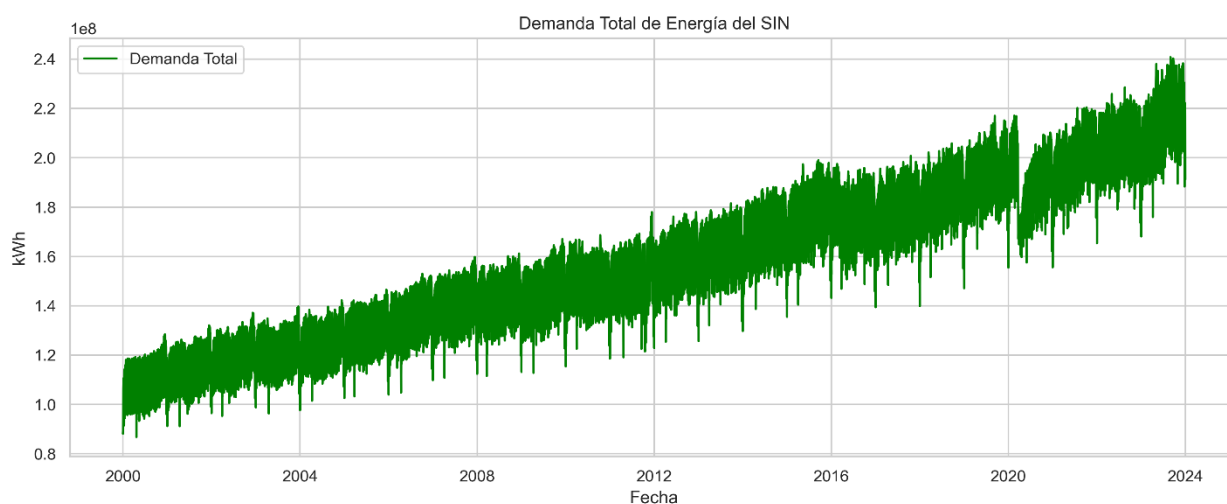


### ***Comportamiento de la Demanda Energética***

La Figura 3 muestra la evolución de la demanda total diaria del Sistema Interconectado Nacional (SIN), evidenciando una tendencia creciente y estacionalidad cíclica marcada por los fines de semana y festivos. Este crecimiento sostenido refleja la expansión económica y el incremento poblacional del país, en concordancia con lo reportado por la Unidad de Planeación Minero Energética (UPME, 2023).

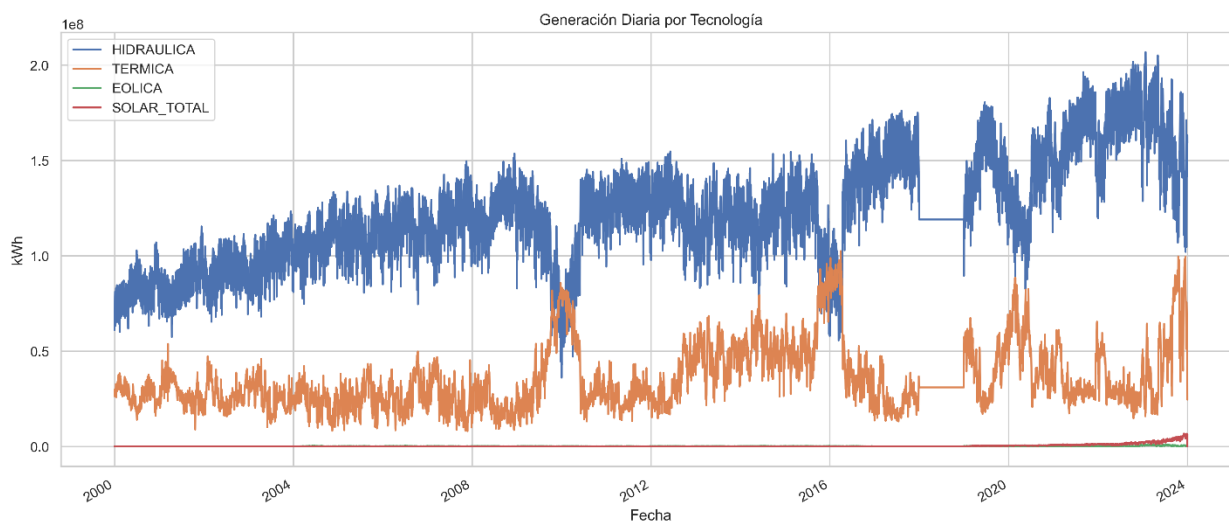
#### **Figura 3**

### ***Comportamiento de la Demanda Energética***

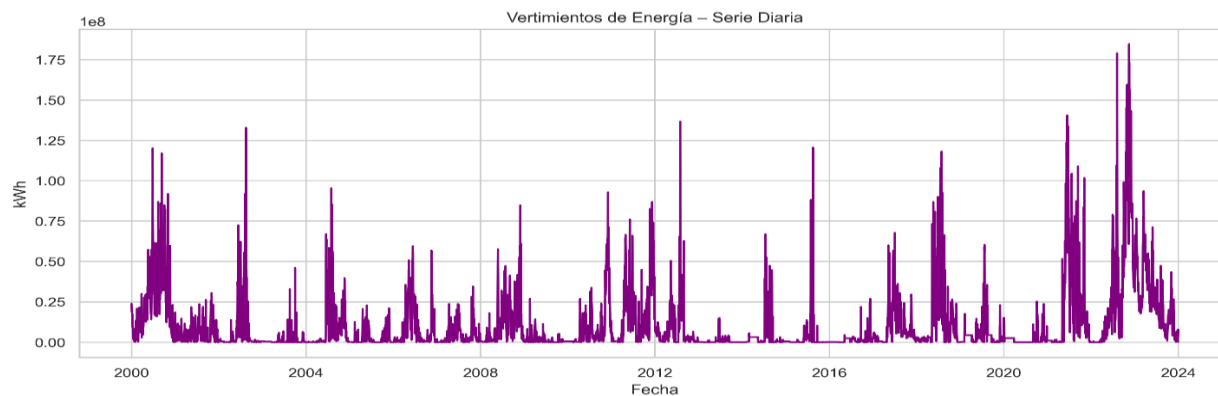


### ***Generación Eléctrica por Tecnología***

El análisis de la Figura 4 revela que la generación hidráulica ha sido históricamente la fuente predominante, aunque su estacionalidad expone al sistema a vulnerabilidades en épocas secas. La generación térmica ha actuado como respaldo en estos períodos críticos, mientras que las tecnologías no convencionales (solar y eólica) muestran una tendencia creciente desde 2019, en línea con la política de transición energética (MinEnergía, 2021).

**Figura 4***Generación Eléctrica por Tecnología**Vertimientos Energéticos*

Se identifican picos significativos de vertimiento energético —principalmente de origen hidráulico— durante periodos de alta hidrología (Figura 5). Estos vertimientos representan una pérdida de oportunidad de generación, indicando restricciones operativas en la red o limitaciones en la capacidad de almacenamiento.

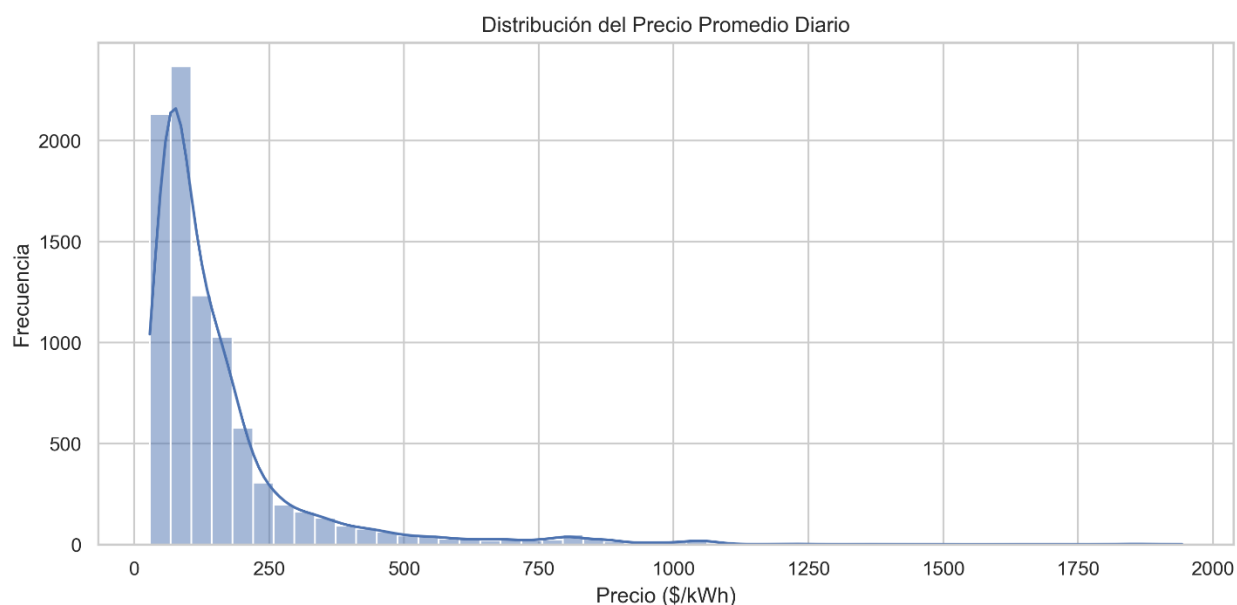
**Figura 5***Vertimientos Energéticos*

### ***Distribución del Precio Promedio Diario***

Como muestra la Figura 6, el precio promedio diario presenta una distribución sesgada a la derecha, con alta concentración en valores bajos (\$50–\$150/kWh) y presencia de colas largas. Esta asimetría justifica la necesidad de aplicar técnicas de normalización para modelado predictivo con redes neuronales, así como transformaciones para cumplir los supuestos de estacionariedad en modelos estadísticos (Hyndman & Athanasopoulos, 2018).

### **Figura 6**

#### *Distribución del Precio Promedio Diario*

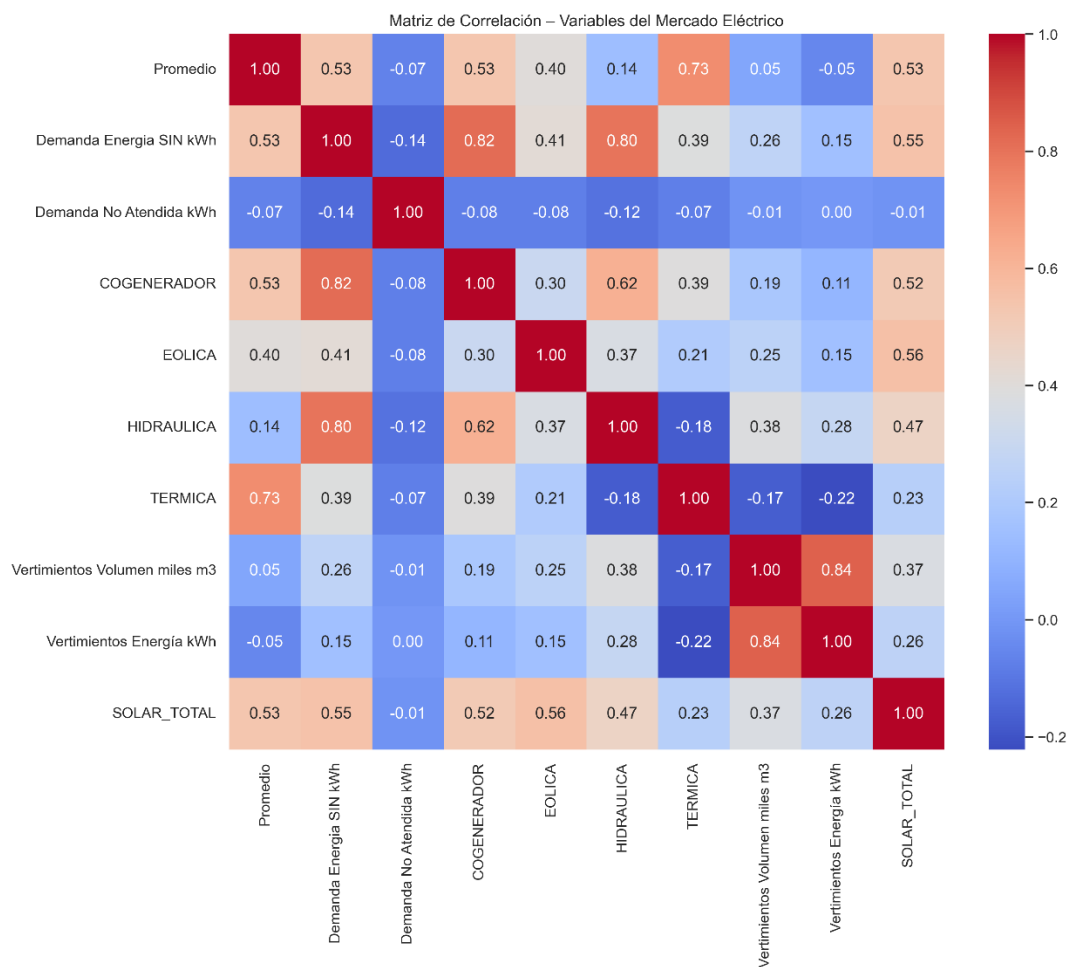


### ***Matriz de Correlación Entre Variables***

La matriz de correlación (Figura 7) evidencia relaciones significativas entre el precio en bolsa y variables como generación térmica ( $r = 0.73$ ), demanda SIN ( $r = 0.53$ ) y generación solar ( $r = 0.53$ ). Estas relaciones sugieren que los precios están fuertemente influenciados por las condiciones de oferta térmica y demanda, reafirmando la conveniencia de incorporar variables exógenas en modelos predictivos multivariados.

Figura 7

## Matriz de Correlación Entre Variables



## **Implementación de Modelos Predictivos**

El objetivo de esta etapa fue entrenar, validar y comparar dos enfoques distintos de modelamiento para la predicción del precio en bolsa de la energía eléctrica en Colombia: un modelo estadístico tradicional basado en series temporales (SARIMAX) y un modelo de aprendizaje profundo (red neuronal LSTM multivariada). A continuación, se describe de manera detallada la implementación de cada uno de estos enfoques.

### **Modelo SARIMAX**

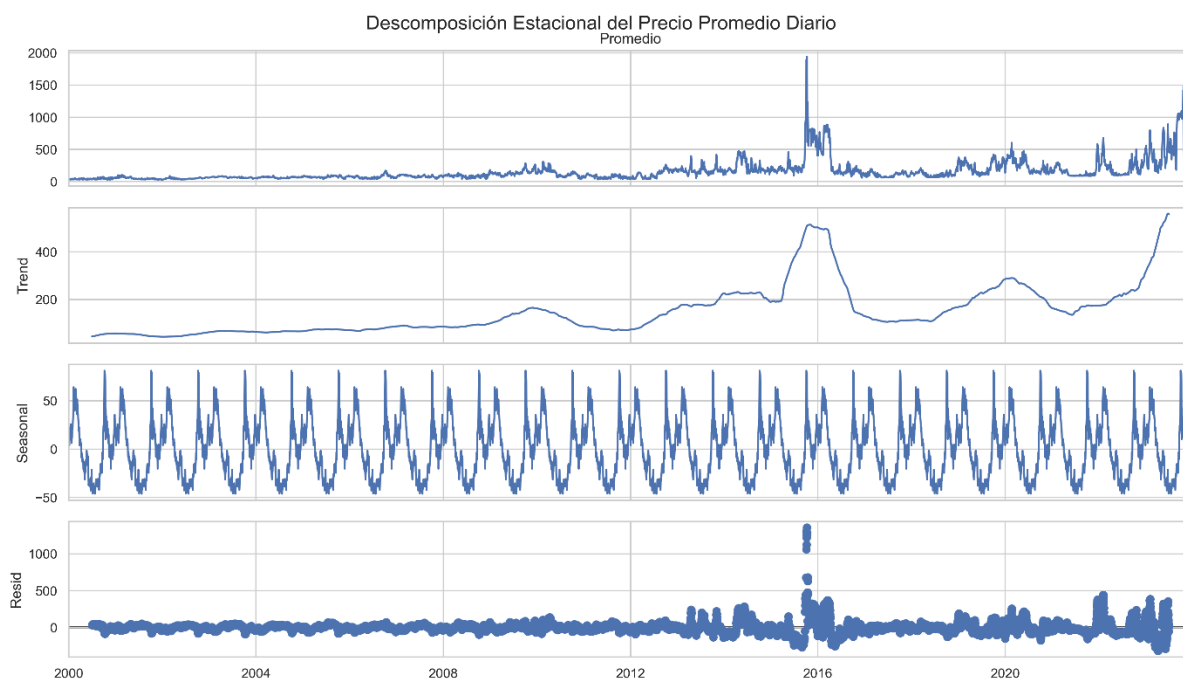
#### ***Descomposición Estacional del Precio Promedio Diario***

Como paso previo al modelamiento estadístico, se realizó la descomposición estacional de la serie temporal del precio promedio diario en bolsa, con el fin de identificar sus componentes estructurales. La descomposición de series temporales permite separar la señal original en tres componentes fundamentales: tendencia (trend), estacionalidad (seasonal) y residuos (resid).

En este estudio se utilizó un modelo aditivo con una periodicidad anual de 365 días, adecuado para datos diarios y consistente con los ciclos hidrológicos del sistema eléctrico colombiano (Hyndman & Athanasopoulos, 2018). Los resultados de esta descomposición se presentan en la Figura 8 y se interpretan a continuación.

## Figura 8

### *Descomposición Estacional del Precio Promedio Diario*



### *Serie Original*

La primera subgráfica representa la serie original del precio promedio diario. Se observan múltiples picos extremos, particularmente durante los periodos 2015–2016, coincidentes con eventos climáticos adversos como el fenómeno de El Niño, los cuales redujeron significativamente los aportes hídricos y forzaron una mayor participación de la generación térmica en el despacho (MinEnergía, 2021). A partir del año 2022 se evidencia una tendencia marcadamente alcista, posiblemente asociada al incremento de los costos operativos, presiones inflacionarias y restricciones estructurales en la oferta de generación.

### *Tendencia (Trend)*

La segunda subgráfica muestra la evolución de la tendencia de largo plazo. Hasta aproximadamente 2013, el comportamiento del precio fue relativamente estable. No obstante, se

identifica un cambio de pendiente entre 2014 y 2016, seguido de un crecimiento sostenido a partir de 2021. Esta componente de tendencia resulta crítica para el ajuste de modelos estadísticos, ya que evidencia la necesidad de aplicar procesos de diferenciación para lograr estacionariedad.

### ***Estacionalidad (Seasonal)***

El tercer panel revela una estacionalidad anual clara, con patrones que se repiten de forma cíclica a lo largo del tiempo. Estos comportamientos estacionales pueden estar asociados a variaciones en la demanda, periodos secos o lluviosos y restricciones operativas del sistema. La presencia de esta estacionalidad justifica el uso de un modelo SARIMAX en lugar de un ARIMA simple, al permitir capturar explícitamente esta variabilidad recurrente (Montgomery et al., 2015).

### ***Residuos (Resid)***

La última subgráfica corresponde a los residuos, es decir, la componente no explicada por la tendencia ni la estacionalidad. Aunque los residuos se concentran alrededor de cero, se observan periodos con alta dispersión y valores extremos, especialmente durante los años 2015–2016 y a partir de 2022. Esto sugiere la influencia de factores exógenos no capturados directamente por la estructura del modelo, como eventos regulatorios, restricciones operativas o crisis energéticas.

En conjunto, la descomposición estacional proporciona una comprensión integral del comportamiento del precio en bolsa y orienta las transformaciones necesarias previas al modelado estadístico. Asimismo, valida empíricamente la pertinencia de utilizar modelos SARIMAX para capturar tanto la estacionalidad como la evolución estructural del precio.

### ***Identificación y Configuración del Modelo SARIMAX***

La identificación de los parámetros del modelo se realizó mediante el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), tanto para los componentes estacionales como no estacionales. Se exploraron distintos valores de los órdenes  $(p, d, q)$  y  $(P, D, Q, s)$ , considerando estacionalidades semanal ( $s = 7$ ) y anual ( $s = 365$ ) en diferentes pruebas exploratorias.

Para garantizar la estacionariedad de la serie, se aplicaron diferenciaciones simples y estacionales, apoyadas en los resultados de la prueba de Dickey-Fuller aumentada (ADF). La implementación del modelo se realizó utilizando la librería statsmodels, que permite el ajuste de modelos SARIMAX mediante máxima verosimilitud.

### ***Ajuste y Validación del Modelo SARIMAX***

La serie fue dividida en conjuntos de entrenamiento y prueba, entrenando el modelo con información histórica hasta el año 2019. El proceso de ajuste incluyó la evaluación de la significancia estadística de los parámetros estimados, el análisis de residuos y la verificación del cumplimiento de supuestos clásicos, tales como ausencia de autocorrelación y comportamiento aproximadamente homocedástico.

El modelo final fue seleccionado con base en criterios de información como AIC y BIC, y preparado para su posterior evaluación comparativa frente al modelo LSTM.

### **Modelo LSTM Multivariado**

#### ***Justificación del Enfoque***

Las redes LSTM (Long Short-Term Memory) son una arquitectura especializada para el tratamiento de series temporales con dependencias de largo plazo y comportamientos no lineales. A diferencia del modelo SARIMA, el enfoque LSTM permite incorporar múltiples variables

exógenas y modelar relaciones complejas sin supuestos estrictos de estacionariedad (Hochreiter & Schmidhuber, 1997; Brownlee, 2021).

Variables Utilizadas:

El modelo LSTM fue alimentado con variables como:

- Precio promedio diario
- Demanda total de energía (SIN)
- Generación por tecnología (hidráulica, térmica, eólica, solar)
- Vertimientos de energía y volumen
- Demanda no atendida

Estas variables fueron seleccionadas con base en su correlación con el precio y su relevancia operativa en el sistema eléctrico colombiano.

### ***Preprocesamiento y Construcción de Secuencias***

Los datos fueron normalizados en el rango  $[0, 1]$  utilizando MinMaxScaler. Luego, se estructuraron en secuencias temporales de entrada de 30 días para predecir el valor del día siguiente. Los conjuntos de datos fueron segmentados en entrenamiento, validación y prueba (70/15/15).

### ***Arquitectura del Modelo***

La red neuronal fue implementada en Keras y configurada con dos capas LSTM de 50 unidades cada una, seguidas de una capa Dropout con tasa de 0.2 para mitigar el sobreajuste. La capa de salida fue una capa densa de una sola neurona, adecuada para problemas de regresión.

La función de pérdida utilizada fue el error cuadrático medio (MSE), y el proceso de optimización se realizó mediante el algoritmo Adam, ampliamente utilizado en modelos de aprendizaje profundo.

### ***Entrenamiento y Evaluación***

El modelo LSTM fue entrenado durante múltiples épocas, incorporando validación por épocas para controlar el sobreajuste. Adicionalmente, se implementó un esquema de evaluación tipo rolling o one-step-ahead, el cual permite simular un entorno operativo real en el que el modelo realiza predicciones diarias incorporando progresivamente nueva información observada.

Este diseño permitió capturar dinámicas no estacionarias del mercado eléctrico colombiano y evaluar la capacidad predictiva del modelo bajo diferentes condiciones históricas.

## **Resultados y Discusión**

En esta sección se presentan los resultados obtenidos a partir de la implementación de los modelos SARIMAX y LSTM para la predicción del precio promedio diario en bolsa del mercado eléctrico colombiano. Ambos enfoques fueron evaluados bajo esquemas de validación diseñados para reflejar condiciones realistas de uso operativo, permitiendo un análisis comparativo tanto cuantitativo como cualitativo.

Las evaluaciones se realizaron considerando series reales y estimadas, métricas de error y comportamiento temporal de las predicciones, con especial atención a la capacidad de los modelos para capturar la alta volatilidad observada en el mercado, particularmente en periodos de estrés hidrológico y cambios estructurales recientes.

### **Criterio General de Evaluación y Comparabilidad Entre Modelos**

Con el fin de garantizar una comparación justa y metodológicamente consistente entre los modelos SARIMAX y LSTM, se adoptaron esquemas de evaluación que, si bien difieren en su implementación computacional, responden a un mismo principio conceptual: evaluar la capacidad predictiva de los modelos bajo un escenario de actualización progresiva de la información disponible, evitando el uso de datos futuros no observables en cada instante de predicción.

En ambos casos, se privilegió un enfoque de tipo rolling forecasting origin, en el cual las predicciones se realizan de manera secuencial a lo largo del período de evaluación, incorporando información real conforme esta se vuelve disponible. No obstante, debido a las diferencias estructurales entre los modelos estadísticos tradicionales y los modelos basados en redes neuronales, fue necesario adaptar el esquema de evaluación a las características computacionales y metodológicas de cada enfoque.

## Resultados del Modelo LSTM

El modelo LSTM fue entrenado utilizando una arquitectura multivariada, con una ventana temporal de 30 días y normalización Min–Max aplicada previamente a las variables de entrada. El entrenamiento se realizó una única vez sobre el conjunto de datos históricos, y la evaluación se llevó a cabo mediante un esquema one-step-ahead, en el cual cada predicción diaria se realizó utilizando la información real más reciente disponible (teacher forcing).

### *Desempeño Cuantitativo*

Bajo este esquema de evaluación, el modelo LSTM obtuvo los siguientes resultados:

- RMSE (escala normalizada): 0.0229
- MAE (escala normalizada): 0.0143

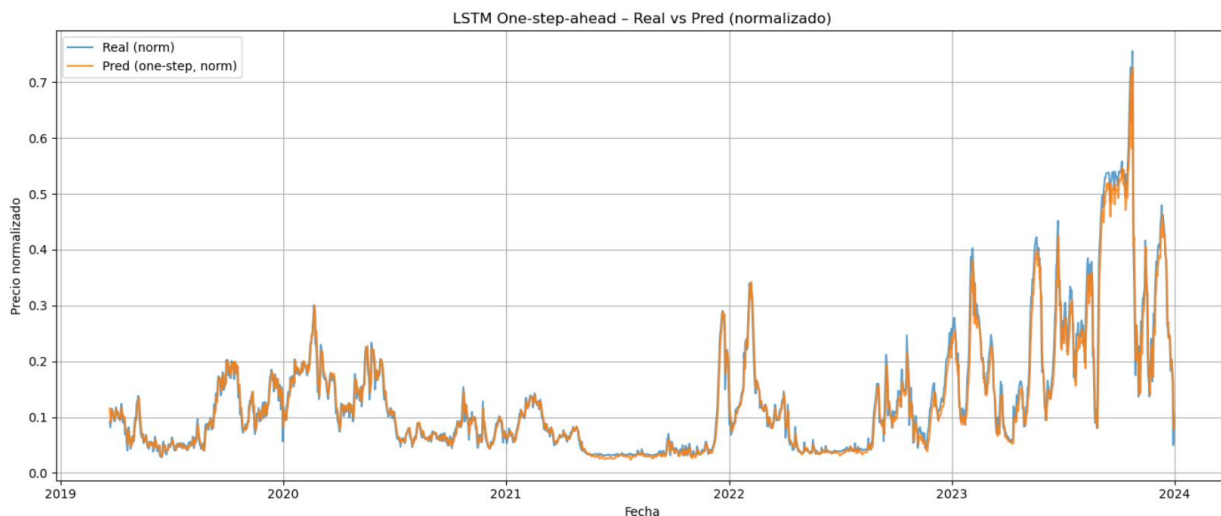
Al desnormalizar las predicciones para llevarlas nuevamente a la escala real del precio en bolsa (\$/kWh), se obtuvieron las siguientes métricas:

- RMSE (\$/kWh): 41.89
- MAE (\$/kWh): 25.32

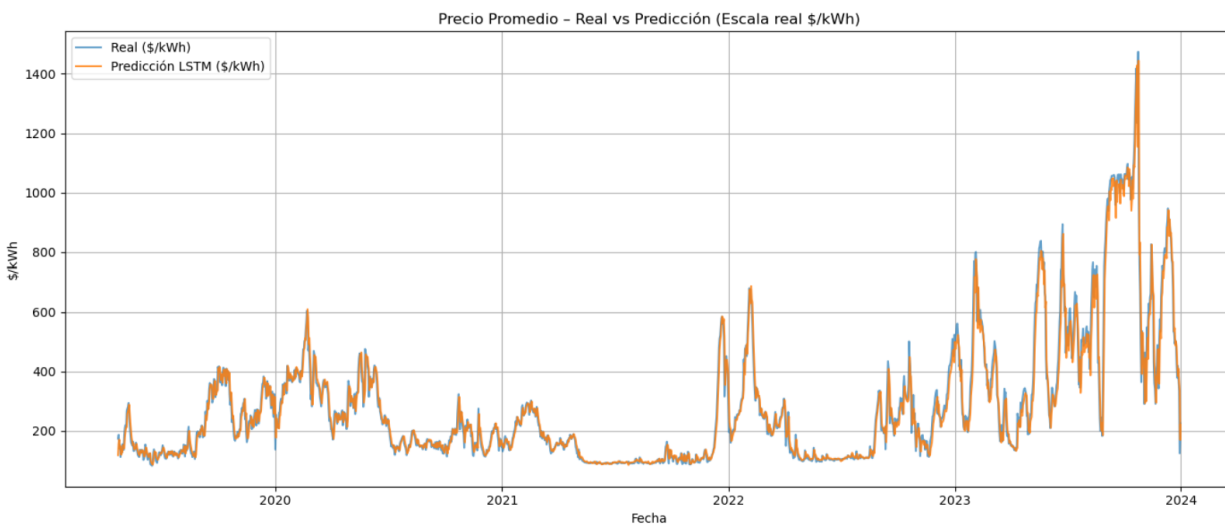
Estos resultados evidencian una alta capacidad del modelo para reproducir el comportamiento diario del precio, con errores relativamente bajos incluso en la escala real, considerando la elevada volatilidad característica del mercado eléctrico colombiano.

**Figura 9**

*Predicción del Modelo LSTM (Normalizado) vs Precio Real en Bolsa*

**Figura 10**

*Predicción del Modelo LSTM (Escala Real) vs Precio Real en Bolsa*



Las gráficas de comparación entre los valores reales y las predicciones muestran un alto grado de ajuste, especialmente en periodos de estabilidad moderada. Durante episodios de alta

volatilidad, el modelo logra seguir adecuadamente la tendencia general, aunque presenta suavización en picos extremos, lo cual es consistente con el comportamiento típico de modelos basados en redes neuronales recurrentes. En las gráficas de la Figura 9 y Figura 10 se pueden observar las predicciones realizadas mediante el modelo LSTM implementado en escala normalizada y real.

### ***Justificación del Esquema One-Step-Ahead para el Modelo LSTM***

En el caso del modelo LSTM multivariado, se implementó un esquema de evaluación one-step-ahead, en el cual el modelo se entrena una única vez utilizando el conjunto de entrenamiento y posteriormente genera predicciones día a día, incorporando en cada paso el valor real observado de la variable objetivo. Este enfoque, comúnmente denominado teacher forcing, permite evaluar la capacidad del modelo para aprender relaciones temporales de corto plazo sin que los errores se acumulen de forma progresiva.

Desde el punto de vista computacional, este esquema resulta viable y eficiente para modelos LSTM, dado que el proceso de entrenamiento es el componente más costoso, mientras que la generación de predicciones secuenciales tiene un costo marginal reducido. En consecuencia, la evaluación one-step-ahead permite obtener métricas de desempeño representativas, estables y comparables, sin incurrir en tiempos de ejecución excesivos.

Adicionalmente, este enfoque es consistente con aplicaciones operativas en las que el modelo se entrena periódicamente y se utiliza para realizar predicciones diarias, actualizándose conforme se dispone de nueva información real.

### **Resultados del Modelo SARIMAX**

El modelo SARIMAX fue evaluado utilizando un esquema de backtesting con reentrenamiento periódico cada 30 días, lo que permite simular un escenario operativo más

realista para modelos estadísticos tradicionales. En este enfoque, el modelo se reentrena al inicio de cada bloque temporal utilizando toda la información histórica disponible hasta ese momento y genera predicciones para el bloque siguiente.

Con el fin de garantizar una comparación justa con el modelo LSTM, se realizaron evaluaciones tanto en escala normalizada como en escala real (\$/kWh).

### ***Desempeño Cuantitativo***

Bajo el esquema de backtesting con refit cada 30 días, el modelo SARIMAX obtuvo:

- RMSE (normalizado): 0.0810
- MAE (normalizado): 0.0474
- Número de observaciones evaluadas: 1754

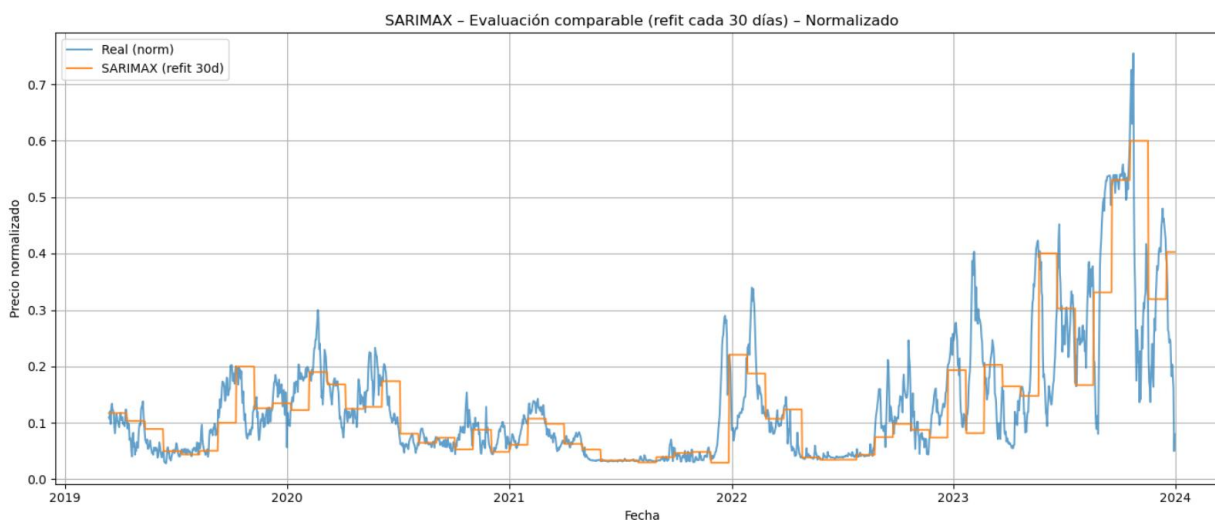
Estos valores indican un desempeño aceptable, aunque inferior al obtenido por el modelo LSTM bajo el esquema one-step-ahead. Al desnormalizar las predicciones, se obtuvieron las siguientes métricas:

- RMSE (\$/kWh): 155.05
- MAE (\$/kWh): 90.78

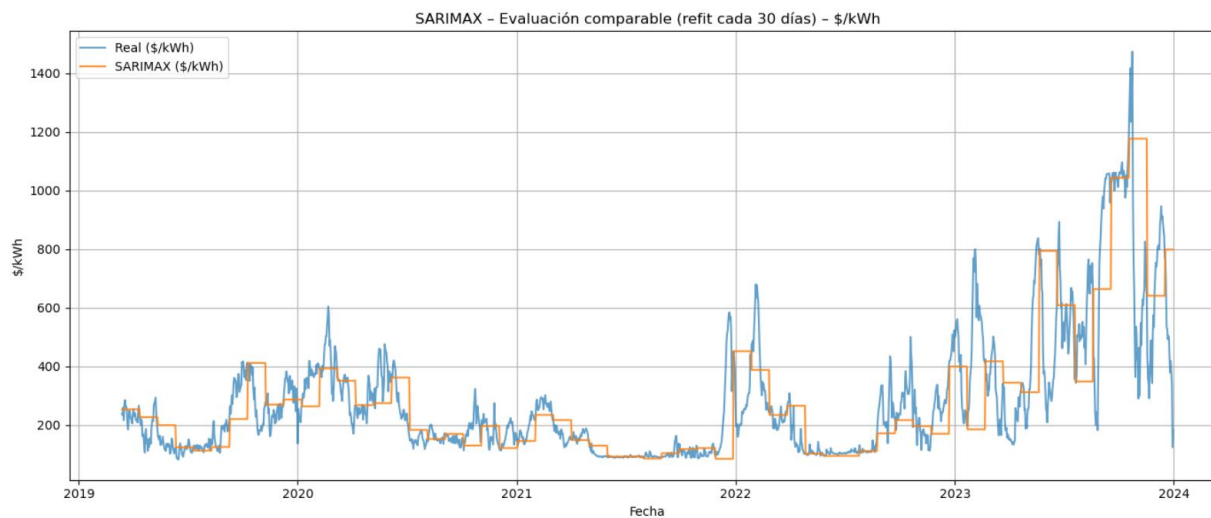
El incremento significativo del error en escala real refleja la dificultad del modelo SARIMAX para adaptarse a cambios abruptos y episodios de alta volatilidad, especialmente en los años más recientes del período analizado.

**Figura 11**

*Predicción del Modelo SARIMAX (Normalizado) vs Precio Real en Bolsa*

**Figura 12**

*Predicción del Modelo SARIMAX (Escala Real) vs Precio Real en Bolsa*



Las gráficas del modelo SARIMAX muestran que el enfoque por bloques genera predicciones con apariencia escalonada, producto del reentrenamiento periódico. Si bien el

modelo captura adecuadamente la tendencia general del precio, presenta rezagos en la respuesta frente a cambios bruscos, particularmente durante eventos extremos.

### ***Justificación del Esquema de Backtesting por Bloques para el Modelo SARIMAX***

Para el modelo SARIMAX, si bien es conceptualmente posible implementar una evaluación one-step-ahead diaria similar a la utilizada en el LSTM, este enfoque resultó computacionalmente inviable en el contexto del presente estudio. Cada actualización diaria del modelo implica la reestimación de sus parámetros mediante máxima verosimilitud, lo cual conlleva un elevado costo computacional, especialmente al considerar la longitud de la serie temporal y la inclusión de componentes estacionales.

En virtud de lo anterior, se adoptó un esquema de backtesting por bloques con reentrenamiento periódico cada 30 días, en el cual el modelo se reestima utilizando toda la información disponible hasta el inicio de cada bloque y se generan predicciones para el horizonte correspondiente. Este enfoque reduce significativamente los tiempos de ejecución, mantiene la coherencia estadística del modelo y representa de manera realista un escenario operativo en el que los modelos estadísticos se recalibran con una periodicidad definida.

Aunque el esquema por bloques introduce una ligera discretización temporal frente al enfoque one-step-ahead, permite capturar adecuadamente los cambios estructurales del precio y evita la acumulación excesiva de error, ofreciendo una evaluación robusta y comparable en términos de desempeño agregado.

### **Comparación Entre SARIMAX y LSTM**

La comparación entre ambos modelos debe interpretarse considerando no solo las métricas de error, sino también el esquema de evaluación utilizado y las implicaciones prácticas de cada enfoque.

**Tabla 2**

*Comparación de Métricas Entre los Modelos Predictivos Implementados.*

Modelo	Esquema de evaluación	RMSE (norm)	MAE (norm)	RMSE (\$/kWh)	MAE (\$/kWh)
LSTM	One-step-ahead	0.0229	0.0143	41.89	25.32
SARIMAX	Backtesting (30d)	0.0810	0.0474	155.05	90.78

*Nota.* La Tabla muestra las métricas obtenidas para cada modelo evaluado

Desde el punto de vista cuantitativo, el modelo LSTM supera de forma consistente al modelo SARIMAX en todas las métricas evaluadas. Esta diferencia se explica principalmente por la capacidad del LSTM para capturar relaciones no lineales y adaptarse dinámicamente a patrones complejos del mercado.

### ***Coherencia Metodológica Entre Ambos Enfoques***

En síntesis, aunque los esquemas de evaluación utilizados para los modelos LSTM y SARIMAX difieren en su implementación específica, ambos responden a un criterio metodológico común orientado a evaluar el desempeño predictivo bajo condiciones realistas de actualización de la información. El uso de one-step-ahead en el LSTM y de backtesting por bloques en el SARIMAX no constituye una inconsistencia metodológica, sino una adaptación necesaria a la naturaleza de cada modelo, garantizando eficiencia computacional y validez comparativa de los resultados obtenidos.

### **Discusión de Resultados**

Los resultados obtenidos confirman que los modelos basados en aprendizaje profundo, como LSTM, presentan ventajas significativas frente a modelos estadísticos tradicionales en

contextos caracterizados por alta volatilidad y no linealidad, como el mercado eléctrico colombiano.

No obstante, el modelo SARIMAX mantiene valor como herramienta interpretativa y de referencia, especialmente en escenarios donde la transparencia del modelo, el bajo costo computacional y la facilidad de implementación son factores relevantes.

Desde una perspectiva de planeación y toma de decisiones, el modelo LSTM resulta más adecuado para aplicaciones que requieren alta precisión en el corto plazo, mientras que SARIMAX puede ser útil como modelo base o de contraste, así como para análisis exploratorios y validación cruzada de resultados.

En síntesis, ambos modelos fueron implementados y evaluados de manera coherente y comparable. Los resultados muestran que el modelo LSTM ofrece un mejor desempeño predictivo, especialmente bajo esquemas de predicción diaria one-step-ahead, mientras que SARIMAX presenta limitaciones frente a la dinámica reciente del mercado.

Es importante señalar que el presente análisis presenta algunas limitaciones. En particular, el modelo SARIMAX no incorporó variables exógenas explícitas, y la evaluación se centró únicamente en el precio promedio diario, lo cual puede subestimar dinámicas intradiarias relevantes. Adicionalmente, eventos extremos como los observados durante 2022–2023 representan un desafío significativo para cualquier enfoque puramente histórico. Sin embargo, estos hallazgos respaldan la elección de modelos de aprendizaje profundo como herramientas robustas para la predicción de precios en mercados eléctricos altamente volátiles, y sirven como base para trabajos futuros orientados a la integración de variables exógenas y enfoques híbridos.

## Conclusiones

El presente trabajo abordó la problemática de la predicción del precio promedio diario en bolsa del mercado eléctrico colombiano mediante la implementación y comparación de dos enfoques de modelamiento ampliamente utilizados en series temporales: un modelo estadístico tradicional SARIMAX y un modelo de aprendizaje profundo basado en redes neuronales recurrentes tipo LSTM. El análisis se desarrolló considerando un contexto caracterizado por alta volatilidad, cambios estructurales recientes y una fuerte dependencia de condiciones hidrológicas, factores que representan un desafío significativo para cualquier modelo predictivo.

A partir de los resultados obtenidos, se evidenció que el modelo LSTM multivariado presenta un desempeño superior en términos de precisión predictiva, tanto en escala normalizada como en escala real (\$/kWh), superando de manera consistente al modelo SARIMAX en las métricas de error evaluadas (RMSE y MAE). Esta ventaja se atribuye principalmente a la capacidad del LSTM para capturar relaciones no lineales y patrones temporales complejos, así como a su mayor flexibilidad para adaptarse a cambios abruptos en la dinámica del mercado.

El esquema de evaluación one-step-ahead utilizado para el modelo LSTM permitió una evaluación realista de su capacidad predictiva en escenarios operativos de corto plazo, evitando la acumulación progresiva de error y reflejando de manera adecuada el comportamiento diario del precio. Asimismo, este enfoque resultó computacionalmente eficiente, dado que el costo principal se concentra en la etapa de entrenamiento del modelo, mientras que la generación de predicciones secuenciales presenta un costo marginal reducido.

Por su parte, el modelo SARIMAX, evaluado mediante un esquema de backtesting con reentrenamiento periódico cada 30 días, mostró un desempeño aceptable en la captura de la tendencia general del precio, aunque con limitaciones evidentes frente a episodios de alta

volatilidad y cambios estructurales. Las predicciones generadas bajo este esquema presentaron un comportamiento escalonado y rezagos temporales, lo cual es consistente con la naturaleza de los modelos estadísticos tradicionales y con el enfoque de recalibración por bloques adoptado para reducir el costo computacional.

Es importante resaltar que, si bien los esquemas de evaluación utilizados para ambos modelos difieren en su implementación específica, estos responden a un mismo principio metodológico orientado a evaluar el desempeño predictivo bajo condiciones realistas de actualización de la información disponible. La utilización de enfoques distintos no constituye una inconsistencia metodológica, sino una adaptación necesaria a la naturaleza computacional y estructural de cada modelo, garantizando la validez y comparabilidad de los resultados obtenidos.

Desde una perspectiva práctica, los resultados sugieren que el modelo LSTM resulta más adecuado para aplicaciones que requieren alta precisión en la predicción diaria del precio, tales como apoyo a la toma de decisiones operativas, análisis de corto plazo y estrategias de gestión del riesgo. En contraste, el modelo SARIMAX mantiene valor como herramienta base de referencia, debido a su interpretabilidad, menor complejidad y facilidad de implementación, especialmente en contextos donde la transparencia del modelo es un factor relevante.

En conclusión, los hallazgos de este estudio respaldan el uso de modelos de aprendizaje profundo como herramientas robustas y eficaces para la predicción de precios en mercados eléctricos altamente volátiles. No obstante, también evidencian la importancia de seleccionar esquemas de evaluación y estrategias de implementación acordes con la naturaleza de cada modelo, así como de considerar enfoques complementarios que integren las fortalezas de ambos paradigmas en futuros desarrollos.

## Recomendaciones y Trabajos Futuros

A partir de los resultados obtenidos y del análisis comparativo desarrollado en el presente trabajo, se identifican diversas oportunidades de mejora y líneas de investigación futura orientadas a fortalecer la capacidad predictiva de los modelos y ampliar su aplicabilidad en contextos reales del mercado eléctrico colombiano.

En primer lugar, se recomienda explorar la incorporación explícita de variables exógenas en ambos enfoques de modelamiento. En el caso del modelo SARIMAX, la inclusión de variables exógenas adicionales relacionadas con condiciones climáticas (precipitación, temperatura, radiación solar), variables operativas del sistema eléctrico (reserva operativa, indisponibilidades de generación, margen de reserva), así como variables de mercado y regulatorias (precio de escasez, activación del cargo por confiabilidad y cambios normativos) permitiría capturar de manera más integral la dinámica del mercado y mejorar la capacidad de anticipación del modelo frente a eventos extremos y cambios estructurales. De manera análoga, para el modelo LSTM, la integración de información climática y operativa adicional permitiría enriquecer la representación multivariada del sistema y potencialmente reducir los errores de predicción en escenarios extremos.

Como segunda línea de trabajo, se sugiere evaluar enfoques híbridos que combinen modelos estadísticos y modelos de aprendizaje profundo. Por ejemplo, el uso de modelos SARIMAX para capturar componentes lineales y estacionales, junto con redes LSTM encargadas de modelar las dinámicas no lineales residuales, podría ofrecer un balance entre interpretabilidad y capacidad predictiva, especialmente en mercados caracterizados por alta complejidad y variabilidad.

Adicionalmente, resulta pertinente analizar la extensión del horizonte temporal de predicción y la aplicación de los modelos a frecuencias diferentes, tales como precios horarios o intradiarios. Este tipo de análisis permitiría evaluar la escalabilidad de los enfoques propuestos y su utilidad en aplicaciones operativas más exigentes, como la programación de generación y la gestión de portafolios de energía en mercados de corto plazo.

Desde el punto de vista metodológico, futuras investigaciones podrían profundizar en la evaluación de distintos esquemas de validación y actualización de modelos, incluyendo estrategias adaptativas de reentrenamiento que ajusten dinámicamente la frecuencia de recalibración en función de la volatilidad observada del mercado. Este enfoque permitiría optimizar el balance entre precisión predictiva y costo computacional, particularmente en modelos estadísticos tradicionales.

Finalmente, se recomienda considerar el análisis del impacto de eventos extremos, como fenómenos de El Niño o crisis energéticas, mediante enfoques de modelamiento robusto o técnicas de detección de cambios estructurales. Este tipo de análisis resulta especialmente relevante para mercados eléctricos como el colombiano, donde la dependencia de factores climáticos introduce desafíos adicionales para la predicción de precios.

En conjunto, estas líneas de trabajo futuro no solo permitirían mejorar el desempeño de los modelos evaluados, sino también ampliar su utilidad práctica en procesos de planeación, regulación y toma de decisiones estratégicas dentro del sector eléctrico.

### Referencias Bibliográficas

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis: Forecasting and control (5th ed.). John Wiley & Sons.
- Gómez-Cano, L., Catalina-Cuellar, S., & Méndez-Vargas, R. (2021). Modelo de pronóstico para estimar el comportamiento del precio en bolsa de la energía en Colombia. ResearchGate. [https://www.researchgate.net/publication/350859401\\_Modelo\\_de\\_pronostico\\_para\\_estimar\\_el\\_comportamiento\\_delPrecio\\_en\\_bolsa\\_de\\_la\\_energia\\_en\\_Colombia](https://www.researchgate.net/publication/350859401_Modelo_de_pronostico_para_estimar_el_comportamiento_delPrecio_en_bolsa_de_la_energia_en_Colombia)
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Kılıç, D. K., Nielsen, P., & Thibbotuwawa, A. (2024). Intraday electricity price forecasting via LSTM and trading strategy. *Energies*, 17(12), 2909. <https://www.mdpi.com/1996-1073/17/12/2909>
- Lago, J., Marcjasz, G., & De Schutter, B. (2021). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 293, 116983. <https://doi.org/10.1016/j.apenergy.2021.116983>
- Li, W., & Becker, D. M. (2021). Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under

- consideration of market coupling. *Energy*, 235, 121354.  
<https://www.sciencedirect.com/science/article/abs/pii/S0360544221017916>
- Marcjasz, G., Narajewski, M., Weron, R., & Ziel, F. (2022). Distributional neural networks for electricity price forecasting. *Energy Economics*, 110, 106001.  
<https://www.sciencedirect.com/science/article/pii/S0140988321005879>
- Mishra, B. K., Preniqi, V., Thakker, D., & Feigl, E. (2024). Machine learning and deep learning prediction models for time-series: A comparative analytical study for the use case of the UK short-term electricity price prediction. *Discover Energy*, 4(1), 75.  
<https://link.springer.com/article/10.1007/s43926-024-00075-4>
- Muñoz-Santiago, A., Urquijo-Vanstrahlengs, J., Castro-Otero, A., & Lombana, J. (2017). Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH. *Revista Universidad del Rosario*.  
[https://revistas.urosario.edu.co/xml/5095/509554391005/509554391005\\_visior\\_jats.pdf](https://revistas.urosario.edu.co/xml/5095/509554391005/509554391005_visior_jats.pdf)
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- UPME. (2022). Plan energético nacional 2020–2050. Unidad de Planeación Minero Energética.  
<https://www1.upme.gov.co>
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030–1081.  
<https://doi.org/10.1016/j.ijforecast.2014.08.008>
- XM S.A. E.S.P. (2023). Informe de operación del Sistema Interconectado Nacional.  
<https://www.xm.com.co/Paginas/Informes/InformesOperacionSIN.aspx>

Zhao, Z., Wang, C., Nokleby, M., & Miller, C. (2018). Improving short-term electricity price forecasting using day-ahead LMP with ARIMA models. arXiv preprint.

<https://arxiv.org/abs/1801.02485>