



**Universidad
Europea**

UNIVERSIDAD EUROPEA DE MADRID

ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO

MÁSTER UNIVERSITARIO EN ANALISIS DE DATOS MASIVOS (BIG DATA)

TRABAJO FIN DE MÁSTER

**ESTIMACIÓN DE PRECIOS DE ELECTRICIDAD EN EL
MERCADO NO REGULADO COLOMBIANO**

FABIO RUEDA GARCÍA

Dirigido por

Dr. CARLOS WOLFRAM ROZAS RODRIGUEZ

CURSO 2024-2025

TÍTULO: ESTIMACIÓN DE PRECIOS DE ELECTRICIDAD EN EL MERCADO NO REGULADO COLOMBIANO.

AUTOR: FABIO RUEDA GARCÍA

TITULACIÓN: MÁSTER UNIVERSITARIO EN ANALISIS DE DATOS MASIVOS (BIG DATA)

DIRECTOR/ES DEL PROYECTO: Dr. CARLOS WOLFRAM ROZAS RODRIGUEZ

FECHA: SEPTIEMBRE de 2025

RESUMEN

Este proyecto tiene como objetivo desarrollar un modelo predictivo robusto para estimar los precios de la electricidad en el mercado no regulado colombiano, caracterizado por su alta volatilidad y la ausencia de señales de precios claras. La investigación surge ante la necesidad de contar con herramientas analíticas que permitan a los agentes del sector energético tomar decisiones estratégicas informadas, especialmente en un entorno de contratación bilateral sin regulación tarifaria.

El trabajo se desarrolló de forma independiente, pero contó con el valioso apoyo y la tutoría del Dr. Carlos Wolfram Rozas Rodríguez de la Universidad Europea, quien desempeñó el rol de director del proyecto. La investigación se orientó a resolver una problemática real del sector energético colombiano. Para ello, se recopiló y unificó información histórica diaria de precios, demanda, temperatura e indicadores económicos (como el IPC), entre 2006 y 2021. Posteriormente, se entrenaron diversos modelos estadísticos, de aprendizaje automático y redes neuronales profundas, entre ellos ARIMA/SARIMA, Random Forest, XGBoost, SVM, LSTM y GRU.

Los resultados indican que los modelos de Deep Learning, en particular LSTM y GRU, obtuvieron el mejor desempeño general, destacando por su capacidad para capturar la variabilidad horaria y reducir el error de predicción. Se evidenció la importancia de incluir variables exógenas en la mejora del ajuste. El modelo propuesto representa una herramienta de valor para comercializadores, consumidores industriales y reguladores del mercado eléctrico colombiano.

Palabras clave: predicción de precios, series temporales, aprendizaje automático, aprendizaje profundo, energía eléctrica, LSTM y GRU.

ABSTRACT

This project aims to develop a robust predictive model to estimate electricity prices in the Colombian non-regulated market, characterized by its high volatility and the absence of clear price signals. The research arose from the need for analytical tools that enable stakeholders in the energy sector to make informed strategic decisions, especially in a bilateral contracting environment without tariff regulation.

The project was conducted independently, but it benefited from the valuable support and supervision of Dr. Carlos Wolfram Rozas Rodríguez from Universidad Europea, who served as the project director. The research addressed a real-world problem in the Colombian energy sector. Historical daily data on prices, demand, temperature, and economic indicators (such as the Consumer Price Index - CPI) were collected and unified for the period 2006 to 2021. Subsequently, various statistical models, machine learning algorithms, and deep learning neural networks were trained, including ARIMA/SARIMA, Random Forest, XGBoost, SVM, LSTM, and GRU.

The results show that deep learning models, particularly LSTM and GRU, achieved the best overall performance, especially in capturing hourly variability and reducing prediction errors. The importance of incorporating exogenous variables was also highlighted for improving model accuracy. The proposed model serves as a valuable tool for electricity marketers, industrial consumers, and regulatory bodies in the Colombian power market.

Keywords: price forecasting, time series, machine learning, deep learning, electricity, LSTM and GRU.

AGRADECIMIENTOS

Agradezco principalmente a mi familia y mi pareja quienes, con su apoyo diario e incondicional, me motivaron de gran manera la ejecución del TFM y el Máster en general.

Además, quiero expresar mi más sincero agradecimiento al Dr. Carlos Wolfram Rozas Rodríguez, director de este proyecto y profesor de la Universidad Europea, por su valiosa tutoría, orientación constante y apoyo a lo largo del desarrollo del trabajo. Su experiencia y compromiso fueron fundamentales para avanzar en cada etapa del proyecto con claridad y enfoque.

Extiendo también mi gratitud a los directores, profesores y tutores del Máster en Análisis de Datos (Big Data), de la Universidad Europea, quienes, con sus consejos y enseñanzas, aportaron significativamente a la construcción de los conocimientos que sustentan este trabajo.

A mis compañeros de Máster, gracias por el intercambio de ideas, la colaboración y el espíritu de compañerismo que enriquecieron el proceso de aprendizaje.

Dedicatoria

Con el mayor afecto, dedico la ejecución del TFM y el Máster en Análisis de Datos (Big Data), especialmente a mi familia y mi pareja, quienes son las personas más valiosas, me motivan y apoyan incondicionalmente para lograr mis objetivos.

TABLA RESUMEN

	DATOS
Nombre y apellidos:	Fabio Rueda García
Título del proyecto:	Estimación de precios de electricidad en el mercado no regulado colombiano
Directores del proyecto:	Dr. Carlos Wolfram Rozas Rodríguez
El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:	NO
El proyecto ha implementado un producto: (esta entrada se puede marcar junto a la siguiente)	NO
El proyecto ha consistido en el desarrollo de una investigación o innovación: (esta entrada se puede marcar junto a la anterior)	SI
Objetivo general del proyecto:	Desarrollar un modelo predictivo para estimar los precios de electricidad en el mercado de servicios no regulados de Colombia, que permita anticipar fluctuaciones en los consumos y contribuir a la toma de decisiones estratégicas.

Índice

RESUMEN	3
ABSTRACT	4
TABLA RESUMEN	7
Capítulo 1. RESUMEN DEL PROYECTO	12
1.1 Contexto y justificación.....	12
1.2 Planteamiento del problema	12
1.3 Objetivos del proyecto.....	12
1.4 Resultados obtenidos	12
1.5 Estructura de la memoria.....	12
Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE	13
2.1 Estado del arte	13
2.2 Contexto y justificación.....	15
2.3 Planteamiento del problema	15
Capítulo 3. OBJETIVOS	16
3.1 Objetivos generales	16
3.2 Objetivos específicos	16
3.3 Beneficios del proyecto	16
Capítulo 4. DESARROLLO DEL PROYECTO	17
4.1 Planificación del proyecto	17
4.2 Descripción de la solución, metodologías y herramientas empleadas.....	21
4.3 Recursos requeridos	Error! Bookmark not defined.
4.4 Presupuesto	23
4.5 Viabilidad	23
4.6 Resultados del proyecto	25
Capítulo 5. DISCUSIÓN	46
Capítulo 6. CONCLUSIONES.....	47
6.1 Conclusiones del trabajo.....	47
6.2 Conclusiones personales.....	47

Capítulo 7.	FUTURAS LÍNEAS DE TRABAJO	49
Capítulo 8.	REFERENCIAS	50
Capítulo 9.	ANEXOS	52

Índice de Figuras

Figura 1. Muestra Dataset Unificado	25
Figura 2. Muestra Registros Temperatura en Municipios de Colombia	26
Figura 3. Código en Power Query para obtener Temperatura Promedio Diaria.....	26
Figura 4. Muestra Temperatura Promedio Diaria por Municipios de Colombia	26
Figura 5. Muestra Dataset Unificado con Temperatura Promedio	27
Figura 6. Muestra IPC Mensual 2006-2021	27
Figura 7. Muestra Dataset con Variables Exógenas.....	28
Figura 8. Predicción Promedio Modelo SARIMA	29
Figura 9. Predicción Promedio Modelo Árbol de Decisión.....	31
Figura 10. Predicción Promedio Modelo SVM.....	32
Figura 11. Predicción Promedio Modelo Random Forest	34
Figura 12. Predicción Promedio Modelo XGBoost.....	36
Figura 13. Predicción Promedio Modelo SVR.....	37
Figura 14. Predicción Promedio Modelo MLP	39
Figura 15. Predicción Promedio Modelo LSTM	41
Figura 16. Predicción Promedio Modelo GRU	43

Índice de Tablas

Tabla 1. Características de trabajos anteriores.----- 15

Tabla 2. Cronograma del Proyecto. ----- 17

Tabla 3. Comparación Modelos Candidatos.----- 19

Tabla 4. Arquitectura de Datos y Modelos utilizados en el proyecto. ----- 21

Tabla 5. Descripción Dataset. ----- 22

Tabla 6. Presupuesto invertido en el proyecto. ----- 23

Tabla 7. Resultados Modelos Predictivos. ----- 45

Capítulo 1. RESUMEN DEL PROYECTO

1.1 Contexto y justificación

En Colombia, el mercado de electricidad no regulado es un componente clave del sector energético, caracterizado por contratos libremente negociados entre grandes consumidores y generadores, lo que genera incertidumbre ante la falta de señales de precio claras y previsibles. En este contexto, el presente proyecto propone el desarrollo de un modelo predictivo robusto que combine inteligencia artificial, análisis funcional y variables exógenas, con el objetivo de apoyar a los actores del sector en la estimación de precios, la formulación de estrategias de contratación y la gestión de riesgos derivados de la volatilidad del mercado.

1.2 Planteamiento del problema

La mayoría de los estudios sobre predicción de precios de electricidad en Colombia se enfocan en escenarios regulados y en variables de demanda y consumo, sin atender adecuadamente las particularidades del mercado no regulado, donde los precios dependen de negociaciones bilaterales y presentan menor control. Además, son escasos los trabajos que integran de manera sistemática variables exógenas con frecuencia mixta, como la climatología, en modelos orientados a este segmento del mercado. Esto evidencia la necesidad de desarrollar un modelo actualizado, preciso y flexible, que responda a las condiciones específicas del mercado no regulado colombiano.

1.3 Objetivos del proyecto

El objetivo principal de este proyecto es desarrollar un modelo predictivo para estimar los precios de electricidad en el mercado de servicios no regulados de Colombia, que permita anticipar fluctuaciones en los consumos y contribuir a la toma de decisiones estratégicas.

1.4 Resultados obtenidos

Los resultados del proyecto evidencian que los modelos de Deep Learning, especialmente LSTM y GRU, ofrecen el mejor desempeño en la estimación de precios de electricidad en el mercado no regulado colombiano, superando a otros modelos tradicionales de Machine Learning. Estos modelos lograron capturar adecuadamente la variabilidad horaria y reducir el error de predicción, particularmente en municipios con alta estabilidad en los datos. Se confirma la utilidad de incluir variables exógenas como temperatura e IPC para mejorar la precisión del modelo.

1.5 Estructura de la memoria

La memoria del TFM, contiene Antecedentes o Estado del Arte con planteamiento del problema. Objetivos generales y específicos. Desarrollo del Proyecto con Planificación, Desarrollo, Metodologías, Presupuesto y Resultados del Proyecto. Discusión, Conclusiones y Futuras Líneas de Trabajo.

Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE

2.1 Estado del arte

La predicción de precios de electricidad en mercados no regulados es una tarea compleja, dada la alta variabilidad, múltiples influencias exógenas y la necesidad de modelos precisos para apoyar decisiones estratégicas. En Colombia, diversas investigaciones han explorado metodologías estadísticas y de inteligencia artificial para abordar esta problemática.

En el trabajo de (Velásquez, Franco, & García, 2009), se aborda la necesidad de contar con herramientas de predicción accesibles a los agentes del mercado eléctrico, dado que los modelos oficiales de la UPME no están disponibles públicamente. El estudio compara tres modelos: Promedio Móvil Integrado Autorregresivo ARIMA, Perceptrón Multicapa (MLP) y Red Neuronal Autorregresiva (ARNN), usando datos mensuales entre 1995 y 2008. Los resultados muestran que el modelo ARNN tuvo mejor precisión en la predicción que ARIMA y MLP, debido a su capacidad para capturar relaciones no lineales en la serie de demanda eléctrica.

Otro estudio de (Velásquez, Franco, & Olaya, 2010), se enfoca en predecir los precios promedio mensuales de contratos despachados en la bolsa de energía, usando Máquinas de Vectores de Soporte (SVM), comparándolas con ARIMA y MLP. Utilizando datos entre 1996 y 2008, la SVM demostró mayor precisión en horizontes de hasta 12 meses, especialmente en presencia de cambios estructurales en la serie de precios, siendo robusta ante fluctuaciones inducidas por factores climáticos como el fenómeno “Niño”.

La investigación de (Barrientos et al., 2012), considera la necesidad de realizar proyecciones de largo plazo del precio de la electricidad en el mercado mayorista colombiano. Se emplean dos enfoques: modelos econométricos (Regresión Múltiple y VEC) y Redes Neuronales Artificiales (RNA). El modelo econométrico encontró que el precio de la energía está determinado significativamente por variables como la demanda, el nivel de embalses y el PIB. Se identificó una relación inversa entre el nivel de embalses y el precio, y una relación directa con la demanda. Por su parte, el modelo RNA mostró un desempeño deficiente cuando incluía demasiadas variables (como PIB), pero mejoró notablemente al modelar el precio únicamente con oferta real, demanda y nivel de embalses.

La predicción con RNA para el periodo 2010–2017 estimó precios promedio entre 100 y 150 \$/kWh, con picos justificados por crecimiento de demanda y rezagos en la expansión de la oferta. El estudio concluye que las RNA pueden ser útiles si se ajustan adecuadamente las variables de entrada.

El problema identificado por (Hurtado, Quintero, & García, 2014), es la falta de herramientas adecuadas para modelar los precios diarios ofertados por los agentes generadores, en un mercado como el colombiano. El estudio aplica Lógica Difusa y Redes Neuronales para estimar estos precios, y los compara con Regresiones Lineales y modelos con Ecuaciones de Regresión Aparentemente No Relacionadas (SURE). Los resultados muestran que los modelos de

inteligencia artificial son parcialmente efectivos para seguir las tendencias, superando en muchos casos las limitaciones de los enfoques econométricos tradicionales.

Finalmente, (Gallón & Barrientos, 2021), presentan un programa innovador que considera las curvas de precios horarios como funciones continuas. Usando datos de 2000 a 2017, implementa el método de Hyndman-Ullah-Shang para realizar predicciones día-adelante y mes-adelante. Se logra capturar patrones estructurales diarios comunes en las curvas de precios, mejorando significativamente las predicciones respecto a puntos de referencia tradicionales. Esta aproximación es útil para actores que deben estimar balances financieros y negociar contratos bilaterales en ausencia de regulación de precios.

A continuación, se indica una tabla comparativa de las principales características de los trabajos consultados:

Artículo	Problema o Necesidad	Técnicas o Métodos	Resultados
Un modelo no lineal para la predicción de la demanda mensual de electricidad en Colombia (Velásquez, Franco & García, 2009)	Falta de herramientas accesibles para predecir demanda mensual de electricidad.	ARIMA, MLP, ARNN	ARNN mostró mayor precisión que ARIMA y MLP en datos no lineales.
Predicción de los precios promedios mensuales de contratos en el mercado mayorista (Velásquez, Franco & Olaya, 2010)	Pronóstico de precios mensuales de contratos en el mercado mayorista.	SVM, ARIMA, MLP	SVM fue más precisa en horizontes de hasta 12 meses y robusta ante fluctuaciones.
Modelo para el pronóstico del precio de la energía eléctrica en Colombia (Barrientos et al., 2012)	Necesidad de proyecciones de largo plazo del precio de electricidad.	Regresión múltiple, VEC, RNA	VEC manejó mejor las relaciones a largo plazo; RNA útil con entradas adecuadas.
Estimación del precio de oferta mediante inteligencia artificial (Hurtado, Quintero & García, 2014)	Dificultad para estimar precios de oferta diaria en un mercado concentrado en una o pocas empresas.	Redes neuronales, lógica difusa, regresiones lineales, SURE	Modelos IA superaron parcialmente los tradicionales al capturar tendencias.

Forecasting the Colombian Electricity Spot Price under a Functional Approach (Gallón & Barrientos, 2020)	Capturar patrones diarios para predicción del precio spot horario.	Análisis funcional de datos (FDA) con componentes principales funcionales	Predicciones precisas día y mes adelante, superando referencias clásicas.
--	--	---	---

Tabla 1. Características de trabajos anteriores.

2.2 Contexto y justificación

El mercado de electricidad no regulado en Colombia representa un componente muy importante del sector energético, donde grandes consumidores y generadores celebran contratos libremente negociados. La ausencia de una señal de precio clara y predecible en este segmento introduce incertidumbre para la toma de decisiones estratégicas y la evaluación de riesgos financieros.

Este proyecto busca diseñar e implementar un modelo predictivo robusto que integre técnicas modernas de inteligencia artificial, análisis funcional y variables exógenas relevantes, con un enfoque específico en los requerimientos del mercado no regulado colombiano. Su aplicación facilitará a los actores del sector anticipar precios, diseñar estrategias de contratación y mitigar riesgos asociados a la volatilidad del mercado.

2.3 Planteamiento del problema

En consultas realizadas sobre la predicción de precios de electricidad en Colombia, la mayoría de los estudios existentes se centran en la demanda y consumos sobre escenarios regulados, dejando en segundo plano las características particulares del mercado no regulado, donde los precios dependen de negociaciones bilaterales y tienen menos control.

Asimismo, pocos de estos trabajos han integrado de forma sistemática variables exógenas con frecuencia mixta (por ejemplo: climatología), dentro de marcos de predicción orientados al mercado no regulado.

Por tanto, se evidencia la necesidad de obtener un modelo actualizado que permita estimar precios de electricidad con mayor precisión y flexibilidad, adaptado a las condiciones y requerimientos del mercado no regulado colombiano, el cual ayude a los consumidores de estos servicios a tomar decisiones fundamentadas y estratégicas para sus negocios.

Capítulo 3. OBJETIVOS

3.1 Objetivos generales

El objetivo general de este proyecto es desarrollar un modelo predictivo para estimar los precios de electricidad en el mercado de servicios no regulados de Colombia, que permita anticipar fluctuaciones en los consumos por demanda, mejorar la eficiencia operacional y contribuir a la toma de decisiones estratégicas de acuerdo con los requerimientos de sus negocios.

3.2 Objetivos específicos

Los objetivos específicos del proyecto son los siguientes:

- Identificar los factores que influyen en la variación de precios de electricidad en el sector no regulado en Colombia.
- Realizar análisis descriptivo y exploratorio de los datos obtenidos.
- Preparar y normalizar los datos para aplicar a modelos de aprendizaje automático.
- Desarrollar modelos predictivos basados en técnicas de análisis y aprendizaje automático.
- Evaluar la precisión y efectividad de los modelos desarrollados a través de métricas de desempeño.
- Interpretar los resultados de las diferentes pruebas realizadas con los modelos programados.
- Desplegar el modelo seleccionado como mejor opción para ejecutar en el proyecto.
- Generar informes técnicos que documenten los resultados obtenidos y proporcionen recomendaciones estratégicas para los actores del sector energético.

3.3 Beneficios del proyecto

La ejecución de este proyecto aportará un beneficio significativo al sector energético colombiano, especialmente en el mercado de servicios no regulados, al proporcionar una herramienta predictiva que facilite la estimación precisa de los precios de electricidad. Al identificar los factores que inciden en su variación, aplicar técnicas de análisis y aprendizaje automático sobre datos debidamente preparados, y evaluar rigurosamente el desempeño de los modelos, se ofrecerá un sistema capaz de anticipar fluctuaciones del mercado. Esto permitirá a generadores, comercializadores y consumidores mejorar la planificación estratégica, optimizar la contratación de energía para obtener beneficios económicos y reducir la exposición al riesgo financiero derivado de la volatilidad del mercado eléctrico.

Capítulo 4. DESARROLLO DEL PROYECTO

En el este capítulo se describe en detalle las fases de implementación del proyecto de estimación de precios de electricidad en el mercado no regulado colombiano. Se inicia con la planificación del proyecto, estableciendo el cronograma de trabajo, recursos involucrados y responsabilidades asignadas. A continuación, se detalla la solución propuesta, incluyendo la metodología aplicada, los modelos seleccionados y las herramientas tecnológicas utilizadas. También se presentan los recursos técnicos y humanos requeridos, el presupuesto estimado y un análisis de viabilidad técnica y económica del proyecto. Finalmente, se expone una síntesis de los principales resultados obtenidos durante la ejecución, lo que permite evaluar el cumplimiento de los objetivos propuestos.

4.1 Planificación del proyecto

Se elabora el cronograma con las actividades que se desarrollan durante la ejecución del proyecto y posteriormente, se detallan los procesos realizados durante cada una de ellas.

Actividad	Responsable	Abr	May	Jun	Jul	Ago
Descripción del problema de negocio	Fabio Rueda García	X				
Revisión bibliográfica y estado del arte	Fabio Rueda García	X				
Recopilación y análisis de datos	Fabio Rueda García		X	X		
Preparación y normalización de los datos	Fabio Rueda García			X	X	
Diseño e implementación de modelos	Fabio Rueda García				X	X
Interpretación de los resultados	Fabio Rueda García					X
Implementación y despliegue del modelo	Fabio Rueda García					X
Elaboración del informe final	Fabio Rueda García					X

Tabla 2. Cronograma del Proyecto.

4.1.1 Definición del problema y objetivos del modelo.

- Descripción clara del problema: identificación del desafío de estimar precios de electricidad en el mercado no regulado.
- Contextualización del mercado: caracterización del entorno regulatorio, funcionamiento de contratos bilaterales, y actores involucrados.
- Establecimiento de objetivos analíticos y predictivos, incluyendo horizonte de predicción, frecuencia, y tipo de salida esperada.

4.1.2 Revisión bibliográfica y del estado del arte.

- Análisis comparativo de técnicas existentes: Modelos estadísticos, técnicas de inteligencia artificial y enfoques funcionales.
- Evaluación de sus ventajas, limitaciones y aplicabilidad al contexto colombiano no regulado.
- Justificación de los métodos seleccionados para este proyecto.

4.1.3 Recolección y análisis exploratorio de datos. (Hyndman & Athanasopoulos, 2021)

- Fuentes de datos: Históricos de precios spot y contratos bilaterales, variables exógenas como la demanda eléctrica, oferta, nivel de embalses, temperaturas, fenómenos climáticos y precios de combustibles.
- Herramientas: Python (pandas, matplotlib, seaborn).
- Análisis exploratorio (EDA): Visualización de series temporales, estadísticos descriptivos, identificación de estacionalidades, tendencias, valores atípicos y análisis de correlación entre variables.

4.1.4 Preparación de datos. (Géron, 2019)

- Limpieza de datos: Tratamiento de valores faltantes y corrección de outliers si afectan el modelo.
- Transformaciones necesarias: Escalado/normalización, generación de retardos (lags), medias móviles y diferencias logarítmicas.
- Codificación de variables categóricas (si aplica).
- División del conjunto de datos: Establecer dataset de entrenamiento, validación y prueba.

4.1.5 Diseño e implementación de modelos predictivos. (Géron, 2019) (Goodfellow, Bengio, & Courville, 2016), (Zhang, Eddy Patuwo, & Hu, 1998), (Chen & Guestrin, 2016).

- Modelos candidatos:
 - Modelos base: ARIMA/SARIMA.
 - Machine learning: Árboles de decisión, Support Vector Machine, Random Forest, Gradient Boosting (XGBoost) y Support Vector Regression.
 - Deep learning: MLP, LSTM y GRU (para series temporales con alta granularidad horaria).

Modelo	Descripción	Entrenamiento	Ventajas
ARIMA / SARIMA	Modelos estadísticos para series temporales, capturan tendencia, estacionalidad y autocorrelación.	Se ajustan parámetros p, d, q mediante validación cruzada o AIC/BIC.	Fácil de interpretar y adecuado para datos univariados estacionarios.
Árboles de decisión	Modelo de clasificación o regresión que divide el espacio en regiones jerárquicas de decisión.	División de datos y ajuste del árbol según reducción de impureza (Gini, entropía).	Sencillo de implementar, interpretabilidad visual, útil como baseline.

Support Vector Machine (SVM)	Modelo supervisado que busca un hiperplano óptimo para separar clases (clasificación) o predecir valores (regresión).	Maximiza el margen entre datos y el hiperplano, se entrena con técnicas de optimización convexa.	Robusto en alta dimensionalidad, útil con datos no lineales.
Random Forest	Ensamble de múltiples árboles de decisión entrenados con bootstrap, promedia sus predicciones.	Entrena muchos árboles con muestras aleatorias y promedia el resultado.	Reducción del sobreajuste, buena generalización.
Extreme Gradient Boosting (XGBoost)	Técnica de boosting que entrena árboles secuencialmente corrigiendo errores del anterior.	Secuencia de árboles entrenados uno tras otro, minimizando la función de pérdida.	Alta precisión, excelente para datos complejos y no lineales.
Support Vector Regression (SVR)	Versión de SVM para problemas de regresión, ajusta una función tolerante a errores.	Ajusta parámetros del kernel y margen tolerado a errores con optimización convexa.	Control de margen de error, útil en problemas con ruido.
MLP (Perceptrón Multicapa)	Red neuronal densa que aprende representaciones complejas de los datos mediante capas ocultas.	Entrenamiento mediante retropropagación con función de pérdida como MSE o CrossEntropy.	Flexible, puede modelar relaciones no lineales complejas.
LSTM (Long Short-Term Memory)	Red neuronal recurrente diseñada para capturar dependencias de largo plazo en secuencias temporales.	Retropropagación a través del tiempo (BPTT), optimiza para series secuenciales.	Captura patrones temporales y secuenciales complejos.
GRU (Gated Recurrent Unit)	Variante de LSTM más eficiente computacionalmente, mantiene memoria a largo plazo sin tanta complejidad.	Igual que LSTM pero con menor número de parámetros; también usa BPTT.	Menor coste computacional que LSTM, pero rendimiento similar.

Tabla 3. Comparación Modelos Candidatos.

- Metodología de validación: Validación cruzada con series temporales (TimeSeriesSplit) y evaluación con conjuntos separados (validación simple).
- La elección de los modelos considerados en esta fase se basó en los siguientes criterios metodológicos y técnicos:
 - **Capacidad de manejo de series temporales y secuencias.**
Se priorizaron modelos con buen desempeño en la predicción de variables temporales como los precios de electricidad, que presentan estacionalidades,

tendencias, y posibles ciclos irregulares. Modelos como ARIMA/SARIMA y LSTM/GRU están específicamente diseñados para este tipo de estructuras.

- **Adaptabilidad a datos multivariantes y exógenos.**

Dado que se busca integrar variables externas (como temperatura), se seleccionaron algoritmos que admiten múltiples entradas y no linealidades, como Random Forest, XGBoost y redes neuronales profundas.

- **Comparabilidad con estudios previos del sector energético.**

Los modelos fueron seleccionados considerando su uso y eficacia en investigaciones previas en mercados eléctricos, particularmente en Colombia, como se evidencia en el estado del arte.

- **Robustez ante ruido y datos incompletos.**

Se valoraron modelos que toleran valores atípicos y ruido estructural en los datos, como XGBoost, Random Forest y Support Vector Regression, lo cual es útil en mercados con alta volatilidad.

- **Flexibilidad y escalabilidad computacional.**

Se eligieron modelos que pueden entrenarse en tiempos razonables y escalarse a entornos más complejos, con posibilidad de integración en pipelines de predicción. Este es el caso de los modelos híbridos y de aprendizaje automático implementados en Python.

- **Complementariedad entre métodos estadísticos y de aprendizaje automático.**

Se incluyeron modelos híbridos para aprovechar la capacidad explicativa de los métodos estadísticos (como ARIMA) y el poder predictivo de los modelos no lineales, logrando así una solución más robusta y generalizable.

4.1.6 Evaluación del desempeño de los modelos. (Hyndman & Athanasopoulos, 2021)

- Métricas de evaluación: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error) y R^2 (coeficiente de determinación).
- Análisis de errores: Comparación de predicción vs valores reales y análisis gráfico de residuales.
- Selección del mejor modelo según desempeño y estabilidad.

4.1.7 Implementación y despliegue del modelo seleccionado.

- Construcción de un pipeline de predicción automatizado.
- Documentación del modelo y su arquitectura.
- Simulación de escenarios y prueba en nuevos datos.
- Posible implementación en entorno reproducible (Jupyter).

4.1.8 Elaboración del informe y recomendaciones.

- Documentación técnica de los modelos implementados y resultados obtenidos.
- Visualización de resultados (gráficas, dashboards).
- Recomendaciones para el uso del modelo en decisiones energéticas (por ejemplo: diseño de contratos, evaluación de riesgo, optimización de compras).

- Identificación de oportunidades futuras: actualización continua, integración con sistemas de gestión energética o plataformas web.

4.2 Descripción de la solución, metodologías y herramientas empleadas

A continuación, se indica una arquitectura de la metodología y herramientas empleadas en el transcurso del trabajo.

Etapa	Componentes	Salida esperada
1. Ingesta de Datos	CSV/API (Demanda), CSV/API (Temperatura), CSV/API (IPC).	Datos cargados desde múltiples fuentes.
2. Limpieza y Unión	Scripts de limpieza (Pandas / PySpark). Unión por municipio y fecha.	Dataset unificado y limpio.
3. Procesamiento de Datos	Transformaciones, agregaciones, normalización, creación de nuevas variables.	Variables listas para entrenar modelos.
4. Modelado	Modelos de Regresión / Series Temporales (ARIMA, Prophet, LSTM, etc.).	Predicciones de demanda por municipio y fecha.
5. Resultados y Predicciones	Base de datos con predicciones, métricas del modelo.	Resultados almacenados para consulta, con métricas de error estables con poca varianza.
6. Visualización	Plotly: dashboards de demanda, clima, IPC.	Informes interactivos para usuarios finales.

Tabla 4. Arquitectura de Datos y Modelos utilizados en el proyecto.

4.3 Recursos Requeridos

El proyecto se realizará con datos libres obtenidos de la web oficial de la empresa XM, especializada en la gestión de sistemas de tiempo real, la administración del mercado de energía mayorista y el desarrollo de soluciones y servicios de energía e información en Colombia. (XM, 2021).

Adicionalmente, se utilizarán datos de variables exógenas como temperatura de fuentes como el IDEAM; así como datos económicos como el IPC de fuentes como el DANE. (IDEAM, 2010), (DANE, 2018).

Este Dataset incluye información con las siguientes variables:

Columna	Descripción
DEPARTAMENTO	Nombre del departamento de Colombia al que pertenece el municipio indicado en la siguiente columna.
MUNICIPIO	Nombre del municipio colombiano al que se refiere el registro.
AGENTE COMERCIALIZADOR	Nombre del agente encargado de comercializar energía.
NIVEL DE TENSIÓN	Nivel de tensión de la red eléctrica asociado al suministro. Por ejemplo: 1 = baja (< 1 kV), 2 = media (1 kV - 57.5 kV) o 3 = alta tensión (> 57.5 kV).
CONCEPTO	Concepto bajo el cual se registran los valores de consumo o costo energético. Por ejemplo: Precio Promedio Diario (\$/kWh) y Demanda Diaria GWh.
DÍA 1 a DÍA 31	Valores diarios de consumo o costo eléctrico correspondientes a cada día del mes. En este caso, son los valores correspondientes a las variables Precio Promedio Diario (\$/kWh) y Demanda Diaria GWh.
RESUMEN MES ETIQUETA	Etiqueta que indica el tipo de valor mensual agregado. Por ejemplo: Precio Promedio Mes (\$/kWh) y Demanda Mes GWh.
RESUMEN MES VALOR	Valor numérico correspondiente al resumen mensual asociado a las etiquetas anteriores.
AÑO	Año calendario al que corresponde el registro.
MES	Nombre del mes al que corresponde el registro.
ARCHIVO	Nombre del archivo fuente del que se extrajo la información, útil para trazabilidad y auditoría de datos.
TEMPERATURA PROMEDIO	Temperatura media mensual (en grados Celsius), correspondiente al municipio y mes de cada fila, obtenida de registros históricos meteorológicos.
IPC MENSUAL	Valor del Índice de Precios al Consumidor (IPC) para el mes y año del registro, representando la inflación mensual en Colombia.

Tabla 5. Descripción Dataset.

4.3.1 Arquitectura Tecnológica.

Para la ejecución del proyecto, se requieren dos dispositivos notebook con las siguientes características, teniendo en cuenta que participamos dos personas en la elaboración del trabajo:

- Procesador AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz.
- Memoria RAM 16 GB.
- Disco Duro 512 GB.

Además, se emplean las siguientes aplicaciones y herramientas de software para realizar análisis, cálculos y desplegar los modelos en el transcurso del proyecto.

- **Conexión a Internet:** Mediante navegadores Web (Google Chrome, Mozilla Firefox).
- **Lenguaje:** Python 3.12.
- **Librerías:** pandas, NumPy, scikit-learn, statsmodels, XGBoost, Pytorch/keras/tensorflow, matplotlib/seaborn.
- **Entorno de trabajo:** Jupyter Notebook / Visual Studio Code / Google Colab.
- **Documentación:** Word, Excel y PDF.

4.4 Presupuesto

Tipo de coste	Valor	Comentarios
Horas de trabajo en el proyecto	9.600 €	Presupuesto para dos personas a tiempo completo durante 12 semanas → 2 x 12 semanas x 40 h/semana = 960 horas. A un coste estimado de 10 €/hora, total: 960 x 10 = 9.600 €.
Equipo técnico utilizado	1.800 €	Dos equipos portátiles de gama media con capacidad para cómputo y visualización de resultados (aprox. 900 € cada uno).
Software utilizado	0 €	Se utilizarán herramientas de código abierto: Python, Jupyter, Scikit-Learn, Pytorch, Visual Studio Code.
Estudios e informes	200 €	Adquisición puntual de un artículo o informe especializado de consultoría energética.
Materiales empleados	100 €	Almacenamiento externo, libretas técnicas, papelería especializada para seguimiento físico de tareas.

Tabla 6. Presupuesto invertido en el proyecto.

4.5 Viabilidad

El proyecto presenta un alto grado de viabilidad tanto en términos técnicos como económicos. A continuación, se justifica esta afirmación a través de un análisis de la relación coste-beneficio y de sostenibilidad futura.

4.5.1 Viabilidad Técnica.

El desarrollo del modelo predictivo de precios de electricidad se fundamenta en tecnologías maduras y de acceso libre como Python, Scikit-learn, Pytorch y bibliotecas especializadas para análisis de series temporales. Estas herramientas han sido ampliamente validadas en aplicaciones industriales y científicas, lo que asegura su eficacia y confiabilidad.

El cronograma ha sido diseñado para ejecutarse en 12 semanas, permitiendo una distribución eficiente de las tareas, y todo el proceso puede ser desarrollado con hardware de gama media, sin necesidad de infraestructura de alto coste. Además, el equipo técnico requerido está al alcance de cualquier entorno académico o profesional.

4.5.2 Viabilidad económica (relación coste/beneficio).

El coste total estimado del proyecto es de aproximadamente 11.700 €, de los cuales la mayor parte corresponde a horas de trabajo técnico altamente especializado. Sin embargo, los beneficios esperados —tanto para los actores del sector energético como para el avance del conocimiento científico— superan ampliamente esta inversión:

- Mejora en la precisión de estimaciones de precios no regulados.
- Reducción del riesgo financiero en la contratación de energía.
- Aplicación escalable a otros mercados o variables del sector energético.
- Posibilidad de reutilización del modelo como software de base para plataformas analíticas.

En este sentido, se espera un retorno de valor tanto directo (decisiones estratégicas más informadas), como indirecto (transferencia académica, publicaciones, difusión científica).

4.5.3 Sostenibilidad del resultado.

El modelo desarrollado está concebido bajo principios de replicabilidad y mantenimiento sostenible. Al estar construido con herramientas de código abierto y metodologías documentadas, puede ser actualizado fácilmente con nuevos datos, ajustado a otros mercados o mejorado con versiones más avanzadas de algoritmos.

Además, su potencial para integrarse en entornos de simulación, dashboards o sistemas de soporte a decisiones lo convierte en un activo escalable, que puede mantenerse operativo y relevante a lo largo del tiempo, especialmente en un entorno energético dinámico como el colombiano.

4.6 Resultados del proyecto

Esta sección presenta los principales resultados obtenidos a lo largo del desarrollo del proyecto, tras la implementación y evaluación de diversos modelos estadísticos, de Machine Learning y Deep Learning orientados a la estimación de precios de electricidad en el mercado no regulado colombiano. Se incluyen las métricas de desempeño de cada modelo, así como un análisis comparativo que permite identificar sus fortalezas y limitaciones. Además, se discuten los casos particulares de algunos municipios con comportamientos atípicos, destacando los modelos que lograron mejor precisión en escenarios de alta variabilidad. Estos resultados permiten evaluar la aplicabilidad del modelo propuesto y su valor como herramienta predictiva para los actores del sector energético.

4.6.1 Preprocesamiento Dataset

Se inicia el proceso ETL de los datos, obteniendo los mismos desde la web oficial de XM, empresa especializada en la gestión de sistemas de tiempo real y la administración del mercado de energía mayorista en Colombia. (XM, 2021).

Los datos están disponibles en formato xls con frecuencia mensual desde enero de 2006 a octubre de 2021. Durante proceso de transformación, se modifica el formato de los ficheros de xls a xlsx, se unifican inicialmente los archivos entre los años 2006 y algunos meses del 2016 que tienen una distribución similar de sus variables; posteriormente se unifican los ficheros restantes del año 2016 al 2021 y finalmente, se obtiene un solo fichero con los datos completos.

En el dataset unificado se normalizan los datos correspondientes a las variables: Departamento, Municipio y Agente Comercializados, para obtener los valores en mayúscula y con los valores estándar en los registros. **Ver Anexo 1. Dataset.ipynb y Anexo 2. dataset_unificado_2006_2021.csv.**

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
DEPARTAMENTO	MUNICIPIO	AGENTE COMERCIALIZADOR	NIVEL DE TENSION	CONCEPTO	DIA 1	DIA 30	RESUMEN MES ETIQUETA	RESUMEN MES VALOR	AÑO	MES	ARCHIVO			
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	240,75	233,55	Precio Promedio Mes (\$/kWh)	240,96095	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00081	0,00002	Demanda Mes GWh	0,02353	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	215,13	213,89	Precio Promedio Mes (\$/kWh)	214,87284	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,03319	0,02514	Demanda Mes GWh	0,88007	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	220,44	220,54	Precio Promedio Mes (\$/kWh)	215,07424	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,00035	0,00445	Demanda Mes GWh	0,08115	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	196,31	195,32	Precio Promedio Mes (\$/kWh)	197,38315	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00201	0,00178	Demanda Mes GWh	0,0687	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	208,07	206,28	Precio Promedio Mes (\$/kWh)	209,16078	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,01726	0,01272	Demanda Mes GWh	0,63465	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	184,58	182,9	Precio Promedio Mes (\$/kWh)	184,56751	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	Demanda Diaria GWh	0,01043	0,0067	Demanda Mes GWh	0,30304	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	213,06	213,06	Precio Promedio Mes (\$/kWh)	213,05995	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,00513	0,00397	Demanda Mes GWh	0,13122	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	230,32	230,18	Precio Promedio Mes (\$/kWh)	224,02713	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00065	0,00036	Demanda Mes GWh	0,01825	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	150,97	151,31	Precio Promedio Mes (\$/kWh)	151,52607	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	Demanda Diaria GWh	0,07754	0,08432	Demanda Mes GWh	2,41372	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	196,91	197,8	Precio Promedio Mes (\$/kWh)	196,6841	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,00422	0,00286	Demanda Mes GWh	0,11241	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Precio Promedio Diario (\$/kWh)	166,02	166,71	Precio Promedio Mes (\$/kWh)	166,27219	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Demanda Diaria GWh	0,11249	0,08517	Demanda Mes GWh	4,17341	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	185,3	184,67	Precio Promedio Mes (\$/kWh)	184,51613	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	Demanda Diaria GWh	0,00152	0,0014	Demanda Mes GWh	0,04398	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	Precio Promedio Diario (\$/kWh)	182,91	182,91	Precio Promedio Mes (\$/kWh)	182,90918	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	Demanda Diaria GWh	0,00286	0,00279	Demanda Mes GWh	0,08827	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	194,08	194,16	Precio Promedio Mes (\$/kWh)	194,3705	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,06524	0,05589	Demanda Mes GWh	1,82612	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Precio Promedio Diario (\$/kWh)	163,48	163,45	Precio Promedio Mes (\$/kWh)	163,48795	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Demanda Diaria GWh	0,29207	0,20546	Demanda Mes GWh	8,31958	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	Precio Promedio Diario (\$/kWh)	142,3	142,3	Precio Promedio Mes (\$/kWh)	142,3	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	Demanda Diaria GWh	0,05453	0,05494	Demanda Mes GWh	1,95452	2006	ABRIL	Abril_06.xlsx			
ANTIOQUIA	BELLO	ISAGEN S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	168,65	168,65	Precio Promedio Mes (\$/kWh)	168,65905	2006	ABRIL	Abril_06.xlsx			

Figura 1. Muestra Dataset Unificado

A continuación, desde la web de IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales), entidad pública colombiana que genera y suministra información oficial sobre recursos naturales y medio ambiente, se obtienen datos con valores de temperatura de los sensores ubicados en diferentes zonas de Colombia. (IDEAM, 2010).

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5	A ^B _C Column6
1	FechaObservacion	NombreEstacion	Departamento	Municipio	ValorObservado	UnidadMedida
2	01/21/2020 11:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	25	°C
3	01/21/2020 02:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	31.1	°C
4	01/21/2020 04:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	31.2	°C
5	01/21/2020 12:00:00 AM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	23.4	°C
6	01/21/2020 10:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	24.8	°C
7	01/21/2020 08:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	26.2	°C
8	01/21/2020 03:00:00 AM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	23.4	°C
9	01/21/2020 07:00:00 AM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	24.1	°C
10	01/21/2020 07:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	26.2	°C
11	01/21/2020 07:00:00 AM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	24.1	°C
12	01/21/2020 07:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	26.8	°C
13	01/21/2020 02:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	31.9	°C
14	01/21/2020 05:00:00 PM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	30.1	°C
15	01/21/2020 03:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	31.8	°C
16	01/21/2020 05:00:00 AM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	23.6	°C
17	01/21/2020 01:00:00 AM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	23.5	°C
18	01/21/2020 11:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	24.4	°C
19	01/21/2020 03:00:00 AM	AEROPUERTO VASQUEZ COBO	AMAZONAS	LETICIA	23.5	°C
20	01/21/2020 01:00:00 PM	PTO NARIÑO	AMAZONAS	PUERTO NARIÑO	32	°C

Figura 2. Muestra Registros Temperatura en Municipios de Colombia

Mediante el siguiente análisis y tratamiento de datos en la herramienta Power Query de Excel, se obtiene el valor de temperatura promedio diaria, para cada municipio que se tiene registrado en el dataset unificado.

```

// 1. Cargar el archivo CSV
Origin = Csv.Document(File.Contents("D:\Study\UAcademia\Master Análisis de Datos (Big Data)\2. Trabajo Fin de Máster\Info TFM\Datos RH\Temperatura_Mediente_Met_Altre_20200731_2000.csv"),[Delimiter=";", Column=";", Encoding="65001", QuoteStyle=QuoteStyle.None],
// 2. Promover encabezados
#"Escalados promediados" = Table.PromoteHeaders(Origin, [PromoteAllScalars=true]),
// 3. Convertir tipo con cultura en-US (formato MM/DD/YYYY a MM/DD/YYYY)
#"TipoConvertido" = Table.TransformColumnTypes(#"Escalados promediados", {{"FechaObservacion", type date}, {"NombreEstacion", type text}, {"Departamento", type text}, {"Municipio", type text}, {"ValorObservado", type number}, {"UnidadMedida", type text}}, "en-US"),
// 4. Agregar columna sala con la fecha (sin hora)
#"AgregarFecha" = Table.AddColumn(#"TipoConvertido", "Fecha", each Date.From([FechaObservacion]), type date),
// 5. Agrupar por Departamento, Municipio y Fecha
#"Agrupado" = Table.Group(#"AgregarFecha", {"Departamento", "Municipio", "Fecha"}, {{("TemperaturaPromedio", each Number.Round(List.Average([ValorObservado]), 1), type number)}},
// 6. Convertir el valor del promedio a texto con una decimales
#"FormatoTemperatura" = Table.TransformColumnTypes(#"Agrupado", {{("TemperaturaPromedio", each Number.ToText([",", "F", "es-CO"), type text)}},
// 7. Convertir la fecha a formato dd/MM/yyyy como texto
#"FormatoFecha" = Table.TransformColumnTypes(#"FormatoTemperatura", {{("Fecha", each Date.ToText([",", "dd/MM/yyyy"), type text)}},
// 8. Formato final

```

Figura 3. Código en Power Query para obtener Temperatura Promedio Diaria

	A ^B _C Departamento	A ^B _C Municipio	A ^B _C Fecha	A ^B _C TemperaturaPromedio
1	ANTIOQUIA	VEGACHÍ	22/10/2006	23,20
2	ANTIOQUIA	SANTA ROSA DE OSOS	12/09/2006	13,91
3	ANTIOQUIA	VALDIVIA	15/02/2006	21,51
4	ANTIOQUIA	MEDELLÍN	05/09/2006	20,64
5	ANTIOQUIA	VEGACHÍ	12/09/2006	22,87
6	ANTIOQUIA	ENTRERRIOS	12/09/2006	10,11
7	ANTIOQUIA	SANTA BÁRBARA	05/09/2006	20,25
8	ANTIOQUIA	MEDELLÍN	12/09/2006	20,82
9	ANTIOQUIA	JERICÓ	22/10/2006	15,86
10	ANTIOQUIA	CHIGORODÓ	05/02/2006	26,65
11	ANTIOQUIA	SANTA FE DE ANTIOQUIA	22/10/2006	26,40
12	ANTIOQUIA	VALDIVIA	05/09/2006	22,23
13	ANTIOQUIA	SANTA ROSA DE OSOS	05/09/2006	13,52
14	ANTIOQUIA	CHIGORODÓ	15/02/2006	27,44
15	ANTIOQUIA	SANTA ROSA DE OSOS	15/02/2006	11,20
16	ANTIOQUIA	SANTA ROSA DE OSOS	05/02/2006	12,30
17	ANTIOQUIA	RIONEGRO	05/02/2006	16,35
18	ANTIOQUIA	VEGACHÍ	05/09/2006	23,21
19	ANTIOQUIA	BELLO	05/09/2006	22,63
20	ANTIOQUIA	SANTA BÁRBARA	12/09/2006	20,07

Figura 4. Muestra Temperatura Promedio Diaria por Municipios de Colombia

Se agrega una variable exógena con la temperatura promedio de cada municipio correspondiente a los registros que se tienen en el dataset unificado desde enero de 2006 a octubre de 2021.

DEPARTAMENTO	MUNICIPIO	AGENTE COMERCIALIZADOR	NIVEL DE TENSION	CONCEPTO	DIA 1	DIA 30	RESUMEN MES ETIQUETA	RESUMEN MES VALOR	ANO	MES	ARCHIVO	TEMPERATURA PROMEDIO
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Precio Promedio Diario (\$/kWh)	240,75	233,55	1 Precio Promedio Mes (\$/kWh)	240,9695	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Demanda Diaria GWh	0,00081	0,00002	1 Demanda Mes GWh	0,02353	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	215,13	213,89	2 Precio Promedio Mes (\$/kWh)	214,87284	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Demanda Diaria GWh	0,03319	0,02514	2 Demanda Mes GWh	0,80007	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	220,44	220,54	2 Precio Promedio Mes (\$/kWh)	215,07424	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Demanda Diaria GWh	0,00035	0,00445	2 Demanda Mes GWh	0,08115	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Precio Promedio Diario (\$/kWh)	196,31	195,32	1 Precio Promedio Mes (\$/kWh)	197,38315	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Demanda Diaria GWh	0,00201	0,00178	1 Demanda Mes GWh	0,06687	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	208,07	206,28	2 Precio Promedio Mes (\$/kWh)	209,16078	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	2 Demanda Diaria GWh	0,01726	0,01272	2 Demanda Mes GWh	0,63465	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	3 Precio Promedio Diario (\$/kWh)	184,58	182,9	3 Precio Promedio Mes (\$/kWh)	184,56751	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	3 Demanda Diaria GWh	0,01043	0,0067	3 Demanda Mes GWh	0,30304	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	213,06	213,06	2 Precio Promedio Mes (\$/kWh)	213,05995	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Demanda Diaria GWh	0,00513	0,00307	2 Demanda Mes GWh	0,13122	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Precio Promedio Diario (\$/kWh)	230,32	230,18	1 Precio Promedio Mes (\$/kWh)	224,02713	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	1 Demanda Diaria GWh	0,00065	0,00036	1 Demanda Mes GWh	0,01825	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	3 Precio Promedio Diario (\$/kWh)	150,97	151,31	3 Precio Promedio Mes (\$/kWh)	151,52607	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	3 Demanda Diaria GWh	0,07754	0,08432	3 Demanda Mes GWh	2,41372	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	196,91	197,8	2 Precio Promedio Mes (\$/kWh)	196,6841	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Demanda Diaria GWh	0,00422	0,00386	2 Demanda Mes GWh	0,11241	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	3 Precio Promedio Diario (\$/kWh)	166,02	166,71	3 Precio Promedio Mes (\$/kWh)	166,27219	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	3 Demanda Diaria GWh	0,11249	0,08517	3 Demanda Mes GWh	4,17341	2006	ABRIL	Abril_06.xlsx	20,21
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	185,3	184,67	2 Precio Promedio Mes (\$/kWh)	184,91613	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	2 Demanda Diaria GWh	0,00152	0,0014	2 Demanda Mes GWh	0,04398	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	1 Precio Promedio Diario (\$/kWh)	182,91	182,91	1 Precio Promedio Mes (\$/kWh)	182,90918	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	1 Demanda Diaria GWh	0,00286	0,00279	1 Demanda Mes GWh	0,08927	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Precio Promedio Diario (\$/kWh)	194,08	194,16	2 Precio Promedio Mes (\$/kWh)	194,3705	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	2 Demanda Diaria GWh	0,06524	0,05589	2 Demanda Mes GWh	1,82612	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	3 Precio Promedio Diario (\$/kWh)	163,48	163,45	3 Precio Promedio Mes (\$/kWh)	163,48795	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	3 Demanda Diaria GWh	0,20207	0,20546	3 Demanda Mes GWh	8,31958	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	4 Precio Promedio Diario (\$/kWh)	142,3	142,3	4 Precio Promedio Mes (\$/kWh)	142,3	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	4 Demanda Diaria GWh	0,06543	0,05494	4 Demanda Mes GWh	1,95452	2006	ABRIL	Abril_06.xlsx	21,73
ANTIOQUIA	BELLO	ISAGEN S.A. E.S.P.	3	3 Precio Promedio Diario (\$/kWh)	168,65	168,65	3 Precio Promedio Mes (\$/kWh)	168,65905	2006	ABRIL	Abril_06.xlsx	21,73

Figura 5. Muestra Dataset Unificado con Temperatura Promedio

Seguidamente, se obtienen datos oficiales del IPC (Índice de Precios al Consumidor) mensual desde enero de 2006 a octubre de 2021. Este indicador representa el incremento que tienen diferentes productos y servicios en Colombia, establecido y publicado por el DANE (Departamento Administrativo Nacional de Estadística), entidad colombiana encargada de producir y divulgar información estadística oficial. (DANE, 2018).

A	B	C
MES	ANO	IPC MENSUAL
Enero	2006	0,54
Febrero	2006	0,66
Marzo	2006	0,7
Abril	2006	0,45
Mayo	2006	0,33
Junio	2006	0,3
Julio	2006	0,41
Agosto	2006	0,39
Septiembre	2006	0,29
Octubre	2006	-0,14
Noviembre	2006	0,24
Diciembre	2006	0,23
Enero	2007	0,77
Febrero	2007	1,17
Marzo	2007	1,21
Abril	2007	0,9
Mayo	2007	0,3
Junio	2007	0,12
Julio	2007	0,17
Agosto	2007	-0,13
Septiembre	2007	0,08
Octubre	2007	0,01
Noviembre	2007	0,47
Diciembre	2007	0,49

Figura 6. Muestra IPC Mensual 2006-2021

Con la información anterior, se crea una nueva variable exógena en el dataset unificado, indicando el IPC Mensual a cada registro, teniendo en cuenta el mes y año correspondiente.

Después de realizar tratamiento mencionado, se obtiene el dataset definitivo con los registros desde enero de 2006 hasta octubre de 2021 y las variables exógenas incluidas.

Ver Anexo 3. dataset_unificado_2006_2021_con_temperatura_final_con_ipc.xlsx

A		B		C		D		E		F		G		H		I		J		K		L		M		N		O		P	
DEPARTAMENTO	MUNICIPIO	AGENTE COMERCIALIZADOR	NIVEL DE TENSION	CONCEPTO	DIA 1	DIA 30	RESUMEN MES ETIQUETA	RESUMEN MES VALOR	AÑO	MES	ARCHIVO	TEMPERATURA	PROMEDIO	IPC	MENSUAL																
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	240,75	233,55	Precio Promedio Mes (\$/kWh)	240,96095	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00081	0,00002	Demanda Mes GWh	0,02353	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	215,13	213,89	Precio Promedio Mes (\$/kWh)	214,67284	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	AMAGA	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,03319	0,02514	Demanda Mes GWh	0,88007	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	220,44	220,54	Precio Promedio Mes (\$/kWh)	215,07424	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	ANDES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,00035	0,00445	Demanda Mes GWh	0,08115	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	196,31	195,32	Precio Promedio Mes (\$/kWh)	197,38315	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00201	0,00178	Demanda Mes GWh	0,0687	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	208,07	206,28	Precio Promedio Mes (\$/kWh)	209,16078	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	2	Demanda Diaria GWh	0,01726	0,01272	Demanda Mes GWh	0,63465	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	194,58	192,9	Precio Promedio Mes (\$/kWh)	194,56731	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	3	Demanda Diaria GWh	0,01043	0,0067	Demanda Mes GWh	0,30304	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	213,06	213,06	Precio Promedio Mes (\$/kWh)	213,05995	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	APARTADO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,00513	0,00397	Demanda Mes GWh	0,13122	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Precio Promedio Diario (\$/kWh)	230,52	230,18	Precio Promedio Mes (\$/kWh)	224,02713	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	ARBOLETES	EMPRESA ANTIOQUEÑA DE ENERGIA S.A. E.S.P.	1	Demanda Diaria GWh	0,00685	0,00306	Demanda Mes GWh	0,01825	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	150,97	151,31	Precio Promedio Mes (\$/kWh)	151,52607	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMGESA S.A. E.S.P.	3	Demanda Diaria GWh	0,07754	0,08432	Demanda Mes GWh	2,41372	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	196,91	197,8	Precio Promedio Mes (\$/kWh)	196,6841	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,0422	0,02096	Demanda Mes GWh	0,11241	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Precio Promedio Diario (\$/kWh)	166,02	166,71	Precio Promedio Mes (\$/kWh)	166,27219	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BARBOSA	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Demanda Diaria GWh	0,11249	0,08517	Demanda Mes GWh	4,17341	2006	ABRIL	Abril_06.xlsx	20,21	0,45																		
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	Precio Promedio Diario (\$/kWh)	185,3	184,67	Precio Promedio Mes (\$/kWh)	184,91613	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	COMERCIALIZAR S.A. E.S.P.	2	Demanda Diaria GWh	0,00152	0,0014	Demanda Mes GWh	0,04298	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	Precio Promedio Diario (\$/kWh)	182,91	182,91	Precio Promedio Mes (\$/kWh)	182,90918	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	1	Demanda Diaria GWh	0,00296	0,00279	Demanda Mes GWh	0,08827	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Precio Promedio Diario (\$/kWh)	194,08	194,18	Precio Promedio Mes (\$/kWh)	194,3705	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	2	Demanda Diaria GWh	0,06524	0,05589	Demanda Mes GWh	1,82612	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Precio Promedio Diario (\$/kWh)	163,48	163,45	Precio Promedio Mes (\$/kWh)	163,48795	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	3	Demanda Diaria GWh	0,29207	0,20546	Demanda Mes GWh	8,31956	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	Precio Promedio Diario (\$/kWh)	142,3	142,3	Precio Promedio Mes (\$/kWh)	142,3	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	EMPRESAS PUBLICAS DE MEDELLIN E.S.P.	4	Demanda Diaria GWh	0,05453	0,05454	Demanda Mes GWh	1,95452	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	ISAGEN S.A. E.S.P.	3	Precio Promedio Diario (\$/kWh)	168,65	168,65	Precio Promedio Mes (\$/kWh)	168,63905	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		
ANTIOQUIA	BELLO	ISAGEN S.A. E.S.P.	3	Demanda Diaria GWh	0,00977	0,00932	Demanda Mes GWh	0,06212	2006	ABRIL	Abril_06.xlsx	21,73	0,45																		

Figura 7. Muestra Dataset con Variables Exógenas

4.6.2 Resultados de la Experimentación de los Algoritmos de Predicción.

➤ Experimentación con técnica SARIMAX

Se implementó un modelo SARIMAX (Seasonal ARIMA) con variables exógenas para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, utilizando datos del período 2006–2021. Este modelo es una extensión del modelo ARIMA tradicional, adecuado para series temporales que presentan comportamiento estacional.

Características de la técnica empleada:

Modelo utilizado: SARIMAX de statsmodels.

Parámetros del modelo:

- Orden no estacional: (1, 1, 1)
- Orden estacional: (1, 1, 1)₇ (periodicidad semanal).

Tamaño de ventana: No se utilizaron lags personalizados, el modelo trabaja directamente sobre la serie temporal completa con sus rezagos implícitos.

Frecuencia temporal: Series diarias.

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE y R^2 para cada municipio individualmente y luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R^2
CALI	36.36	44.78	-2.5349
BOGOTA	16.74	20.21	-0.1577
VILLAVICENCIO	30.40	35.07	-2.5006
BUCARAMANGA	25.88	33.02	0.3597
MEDELLIN	15.69	18.54	0.5203
YUMBO	12.82	18.23	0.4730
PALMIRA	12.24	15.03	0.3203
PEREIRA	24.48	31.34	0.0683
BOGOTA D.C	17.58	20.26	-0.6560
PROMEDIO	21.35	26.28	-0.4564

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Azul: Datos de entrenamiento reales.

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio SARIMA.

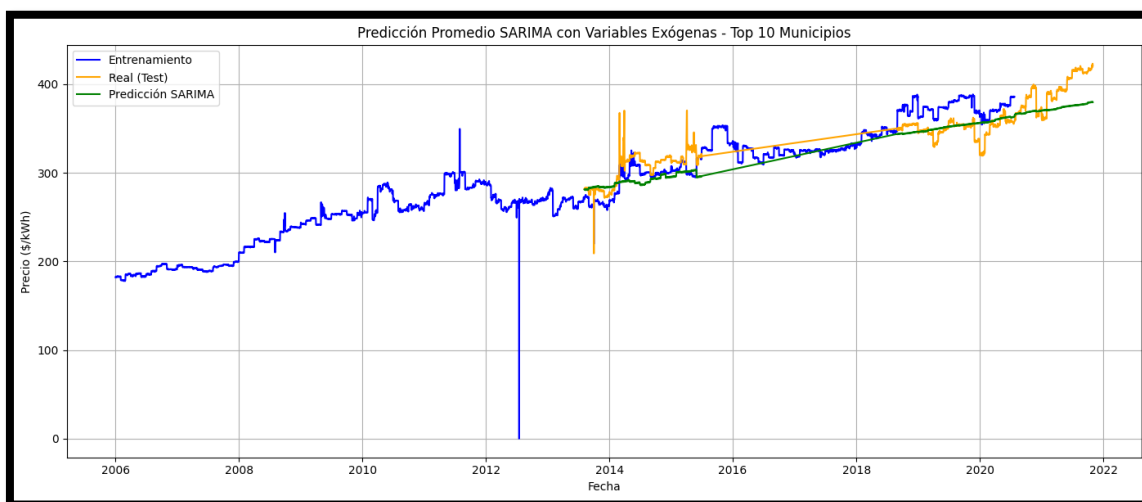


Figura 8. Predicción Promedio Modelo SARIMA

A pesar de incorporar variables exógenas (temperatura e IPC mensual), el modelo SARIMAX sigue presentando un ajuste general bajo. Aunque en algunos municipios como Medellín, Yumbo y Palmira se alcanza un R^2 positivo y aceptable, en otros como Cali, Bogotá y Villavicencio el R^2 es negativo, lo cual ocurre cuando el error cuadrático medio de las predicciones supera la varianza de los datos, lo que refleja una capacidad muy limitada del modelo para capturar la dinámica temporal del precio de la electricidad.

Esto sugiere que el modelo SARIMAX no es suficiente para capturar la complejidad del problema y es necesario explorar modelos no lineales (como Random Forest, XGBoost, MLP, LSTM, etc.).

➤ Experimentación con técnica Árbol de Decisión

Se desarrolló un modelo de regresión basado en Árboles de Decisión para predecir el precio diario de electricidad en los 10 municipios con más registros históricos (2006–2021). Este modelo es adecuado para identificar relaciones no lineales y patrones jerárquicos en los datos.

Características de la técnica empleada:

Modelo utilizado: DecisionTreeRegressor de sklearn.tree.

Parámetros principales: max_depth = 10, random_state = 42.

Tamaño de ventana: 7 días anteriores (LAG_1 a LAG_7) como entrada.

Frecuencia temporal: diaria.

Entrenamiento/Test:

- 80% de los datos para entrenamiento,
- 20% para test.

Variables exógenas incluidas:

- TEMPERATURA PROMEDIO (por municipio).
- IPC MENSUAL (índice de precios por mes).

Municipios utilizados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

MUNICIPIO	MAE	RMSE	R ²
CALI	14.34	38.89	-1.6651
BOGOTA	14.76	18.95	-0.0182
VILLAVICENCIO	5.57	8.33	0.8027
BUCARAMANGA	24.00	38.53	0.1287
MEDELLIN	27.68	35.83	-0.7906
YUMBO	45.16	54.37	-3.6883
BARRANQUILLA	21.81	32.44	0.0662
PALMIRA	22.07	35.72	-2.8339
PEREIRA	83.48	100.30	-8.5458
BOGOTA D.C	27.74	29.96	-2.6292
PROMEDIO	28.66	39.33	-1.9174

Resultados visuales:

Gráfica consolidada con las siguientes curvas:

Azul: Datos reales de entrenamiento.

Naranja: Datos reales de prueba (test).

Verde: Predicción promedio del modelo por día.

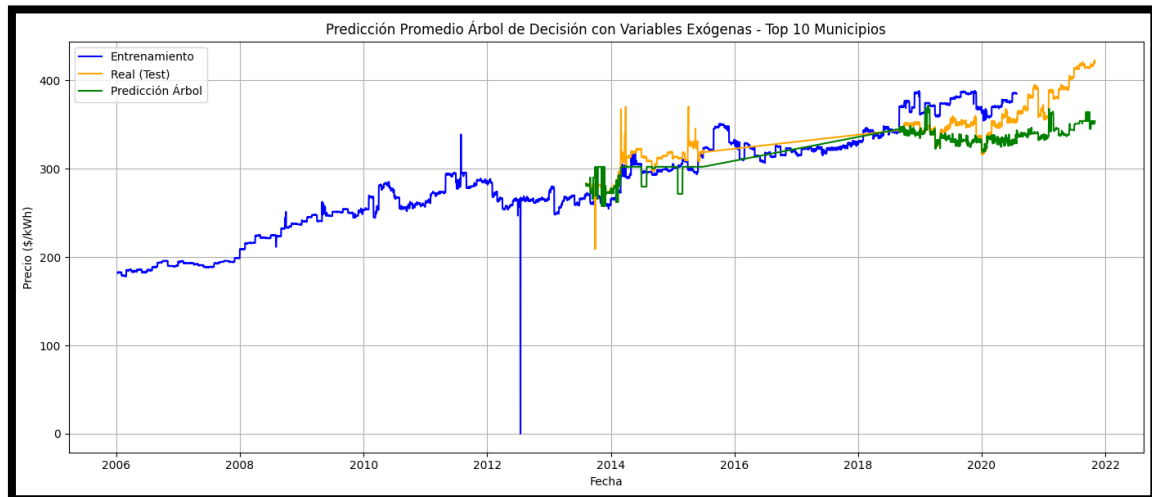


Figura 9. Predicción Promedio Modelo Árbol de Decisión

Aunque en municipios como Bogotá (MAE: 14,76) y Cali (MAE: 14,34) los errores absolutos fueron bajos, en otros como Pereira (MAE: 83,43) el desempeño es muy deficiente. El promedio de R^2 es negativo (-1,9174), lo que indica que el modelo de árboles de decisión no logra generalizar adecuadamente la predicción del precio eléctrico. El bajo desempeño global sugiere que, a pesar de incluir las variables exógenas (Temperatura e IPC), el modelo sufre de sobreajuste o incapacidad de modelar las relaciones temporales complejas, como se evidencia igualmente en la gráfica anterior.

➤ Experimentación con técnica Support Vector Machine (SVM)

Se aplicó un modelo Support Vector Regression (SVR) para predecir el precio diario de electricidad utilizando datos históricos diarios y variables exógenas. Este modelo es efectivo para modelar relaciones no lineales entre variables y es robusto ante ruido en los datos.

Características de la técnica empleada:

Modelo utilizado: SVR de sklearn.svm.

Parámetros utilizados: kernel = 'rbf', C = 100, epsilon = 0.1.

Preprocesamiento: Estandarización de los datos con StandardScaler.

Tamaño de ventana: 7 días anteriores (LAG_1 a LAG_7) como entrada.

Frecuencia temporal: diaria.

Entrenamiento/Test:

- 80% de los datos para entrenamiento.
- 20% para test.

Variables exógenas incluidas:

- TEMPERATURA PROMEDIO (por municipio).

- IPC MENSUAL (índice de precios por mes).

Municipios utilizados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

MUNICIPIO	MAE	RMSE	R ²
CALI	8.99	18.17	0.4185
BOGOTA	7.14	11.93	0.5961
VILLAVICENCIO	3.58	6.17	0.8916
BUCARAMANGA	34.47	64.11	-1.4117
MEDELLIN	29.84	45.58	-1.8980
YUMBO	26.53	44.58	-2.1512
PALMIRA	11.91	20.98	-0.3226
PEREIRA	37.47	56.77	-2.0586
BOGOTA D.C	37.48	41.82	-6.0687
PROMEDIO	21.93	34.46	-1.3338

Resultados visuales:

Gráfico consolidado por día:

Azul: Datos reales de entrenamiento.

Naranja: Datos reales de prueba (test).

Verde: Predicción promedio del modelo SVM.

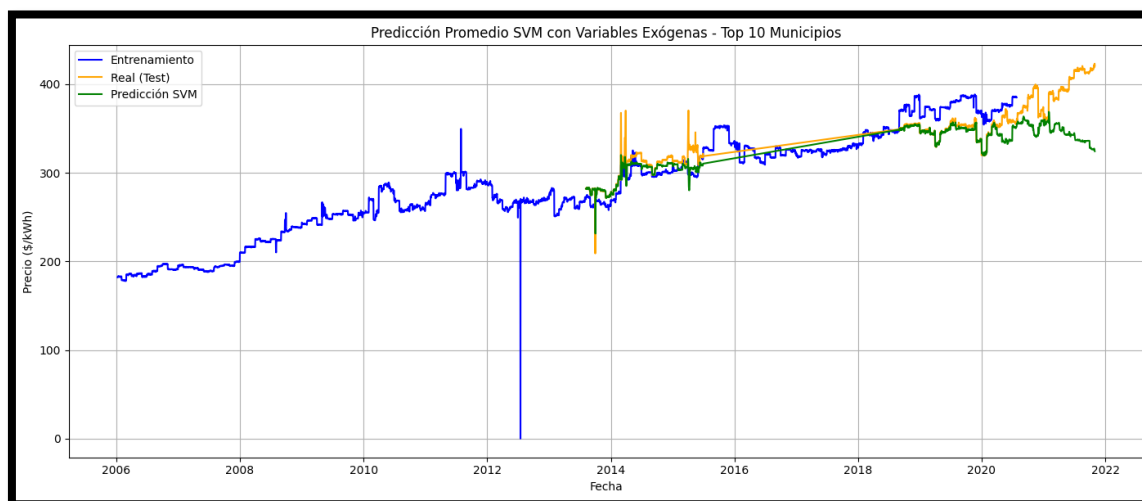


Figura 10. Predicción Promedio Modelo SVM

El modelo SVM logra buenos resultados en municipios específicos (Villavicencio ($R^2 = 0.8916$) y Bogotá ($R^2 = 0.5961$)), pero el rendimiento promedio sigue siendo deficiente. La capacidad de generalización es limitada en presencia de series temporales largas y ruido estructural como valores anómalos. Se recomienda evaluar modelos más robustos o ensamblados como Random Forest o XGBoost.

➤ Experimentación con técnica Random Forest

Se implementó un modelo Random Forest Regressor para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, utilizando datos del período 2006–2021. Este modelo pertenece a la familia de los algoritmos de aprendizaje supervisado basado en ensambles, y permite capturar relaciones no lineales y complejas entre las variables predictoras y la variable objetivo.

Características de la técnica empleada:

Modelo utilizado: RandomForestRegressor de sklearn.ensemble.

Tipo de modelo: Regressor basado en árboles de decisión.

Parámetros: Se utilizaron valores estándar ($n_estimators = 100$, $max_depth = None$, $random_state = 42$).

Variables objetivo: Precio diario promedio de electricidad (\$/kWh).

Lags temporales utilizados: Se aplicó una ventana de 7 días ($lags = 7$) para generar variables auto regresivas que capturan la dinámica temporal local de la serie.

Frecuencia temporal: Series diarias (precio diario histórico para cada municipio entre 2006 y 2021).

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, MAPE y R^2 para cada municipio individualmente, y luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R^2	MAPE
CALI	9.58	23.50	0.0270	2.72
BOGOTA	12.75	16.34	0.2424	4.06
VILLAVICENCIO	4.43	6.71	0.8719	1.18
BUCARAMANGA	23.20	39.49	0.0850	5.37
MEDELLIN	27.56	35.87	-0.7947	6.55
YUMBO	28.37	37.54	-1.2345	7.77
BARRANQUILLA	22.19	34.46	-0.0536	5.94
PALMIRA	11.65	17.53	0.0770	3.23
PEREIRA	30.70	40.25	-0.5375	7.65
BOGOTA D.C	27.96	30.36	-2.7252	6.77
PROMEDIO	19.84	28.20	-0.4042	5.12

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Azul: Datos de entrenamiento reales.

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio del modelo Random Forest.

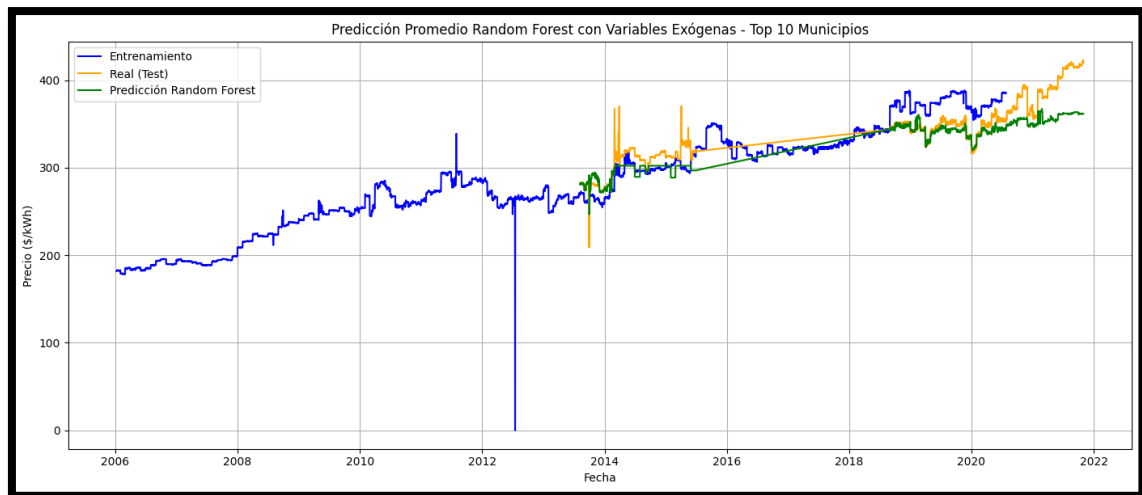


Figura 11. Predicción Promedio Modelo Random Forest

Random Forest logra buenos resultados en algunos municipios como Villavicencio ($R^2 = 0.8719$) o Bogotá ($R^2 = 0.2424$), pero el desempeño global sigue siendo limitado. El MAE y RMSE promedio (19,84 y 28,20) indican que el modelo mejora respecto a árboles de decisión y SVM, pero no supera al modelo SARIMA. La métrica R^2 promedio es negativa (-0,4042), lo cual evidencia que la regresión no supera la predicción con la media, por alta variabilidad de los datos o errores en los mismos que afectan el proceso del algoritmo. El MAPE promedio de 5,12% es razonable, indicando que en términos relativos el modelo tiene un buen comportamiento en algunos municipios.

Se recomienda ajustar hiperparámetros (número de árboles, profundidad máxima) y agregar variables exógenas adicionales o ingeniería de variables para mejorar la generalización.

➤ Experimentación con técnica Extreme Gradient Boosting (XGBoost)

Se implementó un modelo de tipo Gradient Boosting Regressor utilizando la librería xgboost para predecir el precio diario de electricidad en los 10 municipios con mayor número de registros históricos, durante el periodo 2006–2021. Este enfoque permite modelar relaciones no lineales y complejas, aprovechando variables exógenas relevantes.

Características de la técnica empleada:

Modelo utilizado: XGBRegressor de la librería xgboost

Parámetros estándar utilizados: $n_estimators = 100$, $random_state = 42$

Variable objetivo: Precio Promedio Diario (\$/kWh)

Tamaño de ventana temporal: Se aplicó una ventana deslizante de 7 días ($lags = 7$) para capturar dinámicas temporales de corto plazo.

Frecuencia temporal: Series diarias (una predicción por día disponible en los datos históricos).

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, R^2 y MAPE para cada municipio, además de sus promedios generales.

MUNICIPIO	MAE	RMSE	R^2	MAPE
CALI	19.87	40.54	-1.8960	5.38
BOGOTA	17.18	21.49	-0.3103	5.45
VILLAVICENCIO	5.19	7.50	0.8399	1.40
BUCARAMANGA	30.03	49.65	-0.4462	6.98
MEDELLIN	29.03	37.41	-0.9525	6.91
YUMBO	43.68	52.99	-3.4521	12.13
BARRANQUILLA	24.22	37.11	-0.2221	6.49
PALMIRA	8.03	11.64	0.5930	2.25
PEREIRA	50.24	61.22	-2.5567	12.64
BOGOTA D.C	33.32	35.99	-4.2344	8.08
PROMEDIO	26.08	35.55	-1.2637	6.77

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Azul: Datos reales de entrenamiento.

Naranja: Datos reales de prueba (test).

Verde: Predicción promedio del modelo XGBoost.

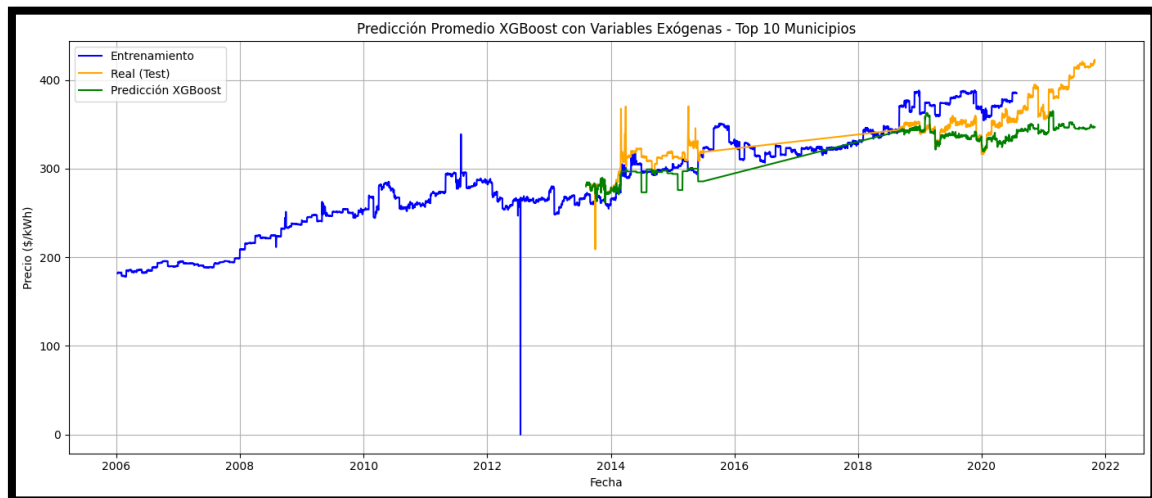


Figura 12. Predicción Promedio Modelo XGBoost

El modelo XGBoost logra buenos resultados en algunos municipios con bajo error absoluto (por ejemplo, Villavicencio con $MAE = 5,19$ y $R^2 = 0,8399$), pero su desempeño general es inestable.

El MAE es menor al de modelos previos como Árboles de Decisión. El MAPE general de 6.77% es competitivo, pero el R^2 negativo refleja que el modelo no mejora sustancialmente la predicción frente a la media. En municipios con alta variabilidad como Pereira y Bogotá D.C. el modelo sufre pérdidas notables de precisión.

En la gráfica se observan algunos datos atípicos que muestran variación en las curvas de test y predicción especialmente.

Se recomienda añadir variables adicionales, o probar enfoques como redes neuronales o modelos híbridos para mejorar la generalización.

➤ Experimentación con técnica Support Vector Regression (SVR)

Se implementó un modelo SVR (Support Vector Regression) con kernel radial (RBF) y parámetros estándar para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, usando datos del periodo 2006–2021. Este enfoque está basado en aprendizaje supervisado no lineal, adecuado para capturar relaciones complejas entre variables.

Características de la técnica empleada:

Modelo utilizado: SVR de sklearn.svm con kernel RBF.

Parámetros: $C = 100$, $\epsilon = 0.1$, $\text{kernel} = 'rbf'$.

Tamaño de ventana: Se aplicó una ventana temporal de 7 días ($\text{lags} = 7$) para generar las características de entrada del modelo.

Frecuencia temporal: Series temporales diarias.

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, R^2 y MAPE por municipio. Luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R^2	MAPE
CALI	8,99	18,17	0,4185	2,37
BOGOTA	7,14	11,93	0,5961	2,24
VILLAVICENCIO	3,58	6,17	0,8916	0,95
BUCARAMANGA	34,47	64,11	-1,4117	7,86
MEDELLIN	29,84	45,58	-1,8980	6,98
YUMBO	26,53	44,58	-2,1512	7,06
PALMIRA	11,91	20,98	-0,3226	3,25
PEREIRA	37,47	56,77	-2,0586	9,14
BOGOTA D.C	37,48	41,82	-6,0687	9,04
PROMEDIO	21,93	34,46	-1,3338	5,43

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Azul: Datos de entrenamiento reales.

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio del modelo SVR.

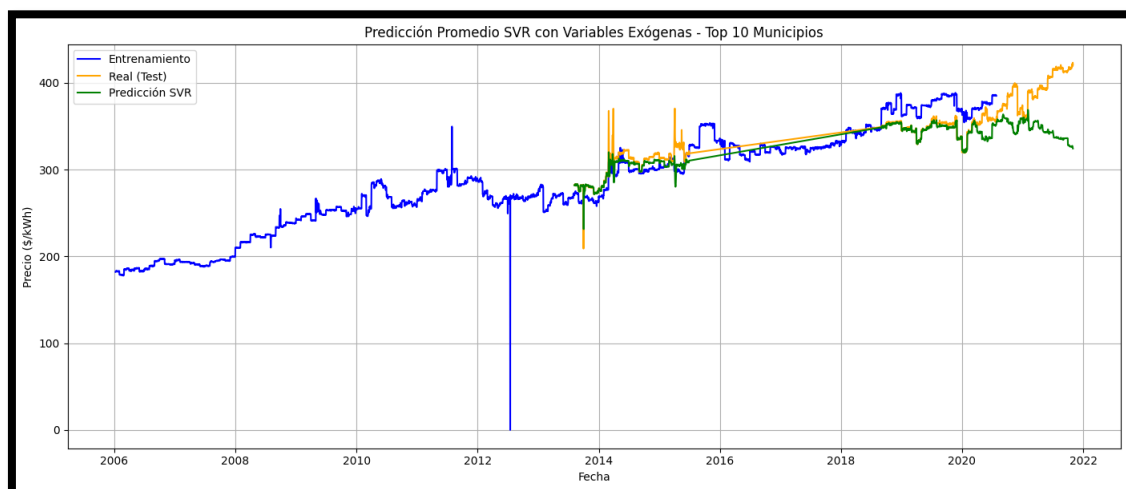


Figura 13. Predicción Promedio Modelo SVR

El modelo SVR con kernel RBF logra muy buenos resultados en municipios con menor variabilidad, como Villavicencio ($R^2 = 0.8916$), Bogotá ($R^2 = 0.5961$) y Cali ($R^2 = 0.4185$). También se destacan sus bajos valores de MAE y MAPE en estos casos. Sin embargo, el modelo pierde precisión y estabilidad en municipios con alta variabilidad, como Bucaramanga, Yumbo, Pereira y Bogotá D.C., donde los valores de R^2 son altamente negativos y el error de predicción es elevado. El promedio general de R^2 negativo (-1.3338) indica que el modelo no supera la predicción basada en la media en la mayoría de los casos. Aun así, los valores medios de MAE y MAPE muestran que el modelo mantiene un buen error absoluto promedio (MAE = 21,93 y MAPE = 5,43%), lo cual es mejor que modelos lineales previos como SARIMA o Árboles de Decisión.

➤ Experimentación con técnica MLP (Perceptrón Multicapa)

Se implementó un modelo MLP (Perceptrón Multicapa) con variables exógenas para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, utilizando datos del periodo 2006–2021. Este modelo utiliza una red neuronal supervisada de tipo feedforward para realizar regresión multivariable.

Características de la técnica empleada:

Modelo utilizado: MLPRegressor de `sklearn.neural_network`.

Estructura de red: Arquitectura multicapa con activación ReLU y función de pérdida cuadrática.

Entrenamiento: Método Optimizador Adam, máximo de 500 iteraciones.

Estandarización: Los datos fueron estandarizados utilizando `StandardScaler` para mejorar la convergencia.

Predicción independiente: Se entrena y predice por separado para cada municipio.

Frecuencia temporal: Series diarias.

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, MAPE y R^2 para cada municipio individualmente, y luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R ²	MAPE
CALI	7,18	12,75	0,7136	1,97
BOGOTA	1,95	5,75	0,9062	0,64
VILLAVICENCIO	3,67	7,80	0,8267	1,01
BUCARAMANGA	4,15	7,72	0,9650	1,04
MEDELLIN	2,88	4,06	0,9770	0,72
YUMBO	2,65	4,22	0,9718	0,74
PALMIRA	6,84	8,45	0,7856	1,93
PEREIRA	4,04	7,41	0,9479	1,07
BOGOTA D.C	56,07	74,53	-21,450	13,40
PROMEDIO	9,94	14,74	-1,60	2,50

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Azul: Datos de entrenamiento reales.

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio MLP.

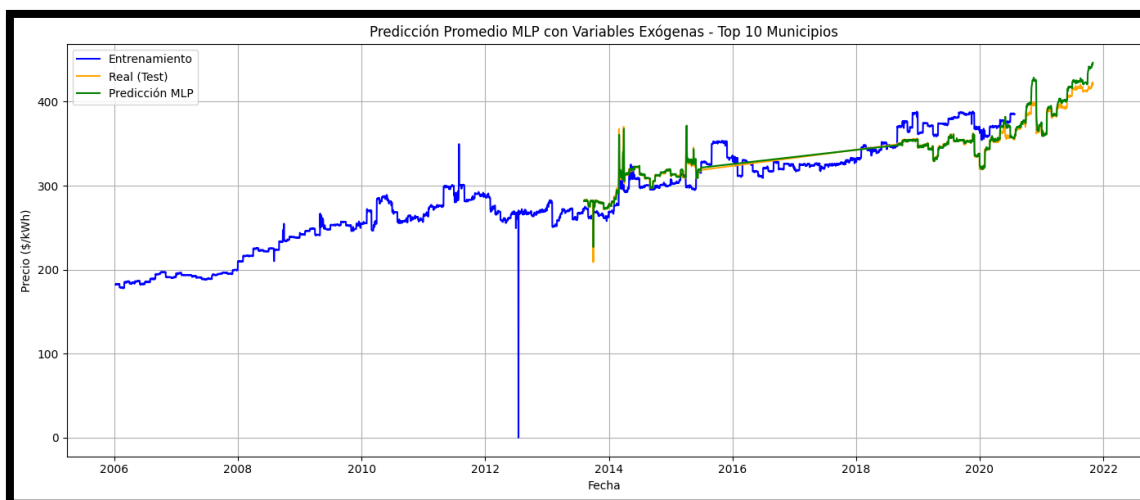


Figura 14. Predicción Promedio Modelo MLP

El modelo MLP con variables exógenas muestra un alto desempeño en la mayoría de los municipios, con R^2 superiores al 0,9, errores bajos (MAE y RMSE) y valores de MAPE menores al 2%, lo que indica buena capacidad de ajuste y generalización. Municipios como Bogotá, Medellín, Bucaramanga y Yumbo presentan resultados excelentes, con errores muy bajos y valores de R^2 muy cercanos a 1. Sin embargo, el modelo presenta una falla crítica en Bogotá D.C., donde los valores extremos y la alta variabilidad de la serie causan una predicción errónea, con un R^2 fuertemente negativo (-21.45) y un MAPE superior al 13%. Esta anomalía distorsiona el promedio general del modelo, provocando un valor de R^2 promedio negativo (-1,60), lo que no refleja el buen desempeño en la mayoría de los municipios.

Se recomienda excluir outliers como Bogotá D.C. en el cálculo del promedio global o aplicar técnicas de ajuste más robustas (como normalización avanzada o redes neuronales más profundas) para regiones con alta variabilidad. También se pueden explorar arquitecturas como LSTM o GRU que procesan mejor las series temporales no lineales.

➤ **Experimentación con técnica LSTM (Long Short-Term Memory)**

Se implementó un modelo LSTM (Long Short-Term Memory) con variables exógenas para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, utilizando datos del período 2006–2021. Este modelo de red neuronal recurrente es adecuado para capturar patrones temporales de largo plazo, siendo eficaz en tareas de predicción secuencial con dependencia temporal.

Características de la técnica empleada:

Modelo utilizado: Red neuronal LSTM implementada con la librería Keras (TensorFlow backend).

Parámetros del modelo:

Arquitectura: Capa LSTM con 50 unidades, activación ReLU.

Capa de salida densa con 1 neurona (predicción de precio).

Optimización: Optimizador Adam (taza de aprendizaje = 0.01)

Función de pérdida: MSE (Error cuadrático medio)

Número de épocas: 50

Preprocesamiento: Escalado con StandardScaler para X e y.

Conversión de fechas de tipo DIA, MES, AÑO a objetos tipo fecha con agrupación diaria.

Tamaño de ventana: Se utilizó una ventana deslizante de 7 días (lags = 7), es decir, para cada predicción se toman los 7 precios anteriores como entrada, junto con las variables exógenas correspondientes.

Frecuencia temporal: Series diarias (valores diarios de precio por municipio).

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, R^2 y MAPE para cada municipio individualmente, y luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R ²	MAPE
CALI	5,40	6,41	0,9276	1,55
BOGOTA	2,65	5,54	0,9129	0,85
VILLAVICENCIO	2,14	4,44	0,9440	0,58
BUCARAMANGA	5,64	9,38	0,9483	1,37
MEDELLIN	2,64	3,82	0,9796	0,66
YUMBO	1,77	2,91	0,9865	0,51
PALMIRA	8,02	10,25	0,6843	2,25
PEREIRA	7,01	8,95	0,9241	1,81
BOGOTA D.C	25,71	29,22	-2,4520	6,26
PROMEDIO	6,78	8,99	0,54	1,76

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio del modelo LSTM.

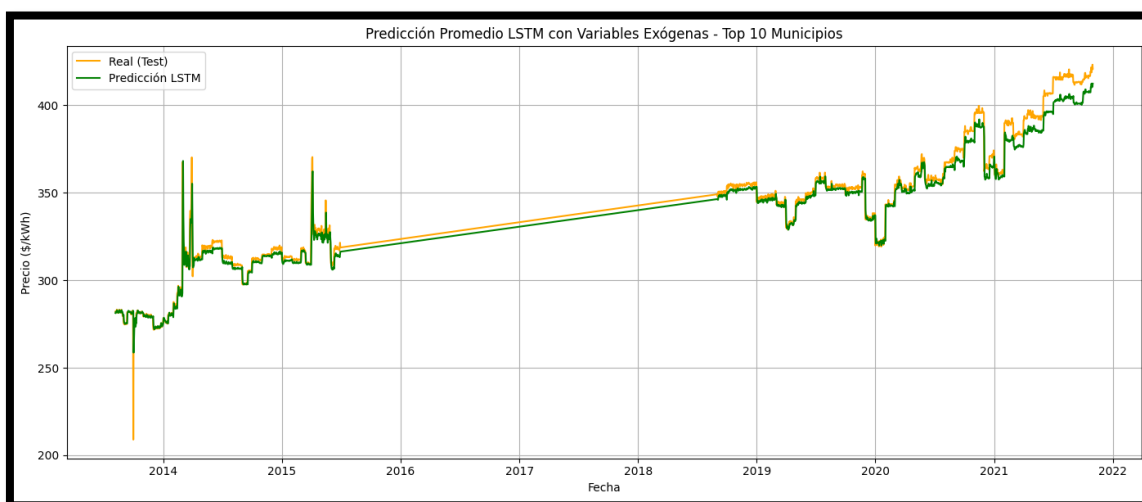


Figura 15. Predicción Promedio Modelo LSTM

Buen desempeño general del modelo LSTM en la mayoría de los municipios, con errores bajos y valores R^2 superiores a 0.9. Municipios como Medellín, Yumbo y Villavicencio muestran predicciones muy precisas. Sin embargo, al igual que en modelos anteriores, el rendimiento en Bogotá D.C. fue inestable y presenta un valor R^2 negativo, lo que indica una regresión peor que la media. Se recomienda revisar este municipio por posibles outliers o reentrenar con mayor regularización. El promedio general muestra un R^2 de 0.54, mejorando notablemente respecto a modelos como MLP, SVR o Random Forest.

➤ Experimentación con técnica GRU (Gated Recurrent Unit)

Se implementó un modelo GRU (Gated Recurrent Unit) con variables exógenas para estimar el precio diario de electricidad en los 10 municipios con mayor cantidad de registros históricos, utilizando datos del período 2006–2021. Este modelo es una red neuronal recurrente optimizada para el aprendizaje secuencial, especialmente útil para series temporales largas con menos complejidad computacional que LSTM.

Características de la técnica empleada:

Modelo utilizado: GRU de tensorflow.keras

Arquitectura:

Capa GRU con 50 neuronas y activación ReLU

Capa de salida densa (1 neurona)

Optimización: Adam (taza de aprendizaje = 0.01)

Función de pérdida: MSE

Épocas: 50

Frecuencia temporal: Series temporales diarias

Segmentación de datos:

- 80% para entrenamiento.
- 20% para validación (test), por municipio.

Variables exógenas utilizadas:

- TEMPERATURA PROMEDIO (por municipio, mes y año).
- IPC MENSUAL (índice económico mensual).

Municipios evaluados: Medellín, Bucaramanga, Yumbo, Cali, Bogotá, Villavicencio, Pereira, Palmira, Barranquilla y Bogotá D.C.

Resultados Métricas de evaluación:

Se calcularon las métricas MAE, RMSE, MAPE y R^2 para cada municipio individualmente y luego se obtuvo el promedio general:

MUNICIPIO	MAE	RMSE	R^2	MAPE
CALI	3,67	5,19	0,9525	1,06
BOGOTA	2,94	5,67	0,9087	0,94
VILLAVICENCIO	2,36	4,56	0,9409	0,64
BUCARAMANGA	2,61	6,57	0,9746	0,69
MEDELLIN	5,91	6,98	0,9321	1,44
YUMBO	7,27	8,26	0,8918	2,05
PALMIRA	7,48	11,02	0,6353	2,08
PEREIRA	6,78	10,47	0,8960	1,75
BOGOTA D.C	36,26	39,79	-5,3983	8,79
PROMEDIO	8,36	10,95	0,19	2,16

Resultados visuales:

Se generó una gráfica consolidada con la evolución del precio promedio diario:

Naranja: Datos de prueba (test) reales.

Verde: Predicción promedio modelo GRU.

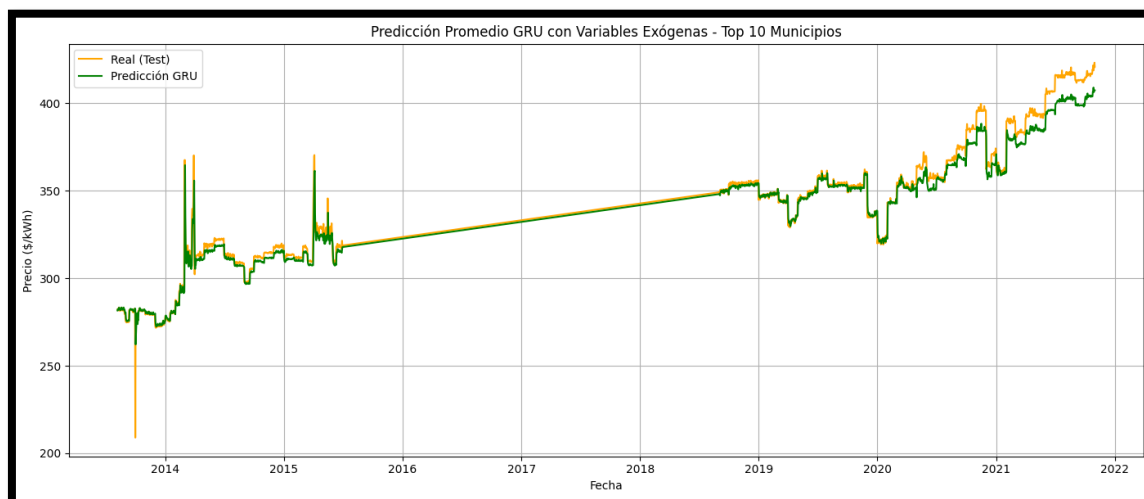


Figura 16. Predicción Promedio Modelo GRU

Buen rendimiento general en municipios como Medellín, Cali y Villavicencio, con valores R^2 mayores a 0,9. El MAE promedio (8,36) y el MAPE promedio (2,16) son bajos, lo que sugiere buena precisión en la predicción. El RMSE promedio es 10,95, también mejor que en varios modelos previos. El R^2 promedio es 0,19, lo que representa una mejora importante respecto a LSTM ($R^2 = 0,54$) y especialmente respecto a modelos como SVM, SVR y Gradient Boosting donde el R^2 fue negativo.

El modelo logra capturar bien la dinámica temporal del precio, y reduce el MAE y el MAPE respecto a modelos como MLP o Random Forest. Sin embargo, presenta inestabilidad en Bogotá D.C. ($R^2 = -5,39$) que afecta significativamente el R^2 promedio.

Se recomienda revisar la calidad de los datos o ajustar la regularización. El rendimiento promedio es competitivo y representa una mejora respecto a otros modelos no recurrentes. La curva de predicción se ajusta adecuadamente a la curva real en la mayoría de municipios, como muestra la última gráfica.

En las visualizaciones de las diferentes técnicas, se observa una línea vertical abrupta en el año 2013 dentro de la curva de entrenamiento. Esta anomalía se debe probablemente a la presencia de datos atípicos, valores nulos o precios extremadamente bajos registrados en ese año para alguno de los municipios.

Por otro lado, las curvas correspondientes al conjunto de prueba (test) y la predicción promedio entre los años 2015 y 2019 muestran una trayectoria más continua y suave.

Esto indica una adecuada cobertura de datos válidos en ese periodo, así como un comportamiento más estable y coherente de los modelos al generar sus predicciones.

Por último, se compartirán los códigos de todos los modelos configurados en esta sección.

Ver Anexo 4. Modelos.ipynb

A continuación, se indica una tabla comparativa con el resumen de resultados de los modelos programados para las diferentes pruebas de predicción.

Modelo	Valor de las métricas (MAE / RMSE / R ² / MAPE)	Conclusión sobre precisión y estabilidad
ARIMA / SARIMA	MAE: 21.35 / RMSE: 26.28 / R ² : -0.4564	Precisión baja. SARIMAX mejora ligeramente el ajuste en algunos municipios, pero el promedio sigue siendo pobre. Se recomienda explorar modelos no lineales y nuevas variables exógenas.
Árboles de decisión	MAE: 28,66 / RMSE: 39,33 / R ² : -1,9174	Precisión deficiente. Aunque el error en algunos municipios es bajo, el ajuste general es muy pobre. No captura bien la dinámica temporal ni mejora significativamente al incluir variables exógenas. Se recomienda probar modelos más complejos como Random Forest, XGBoost o redes neuronales.
Support Vector Machine (SVM)	MAE: 21,93 / RMSE: 34,46 / R ² : -1,3338	Desempeño irregular. Aunque el modelo logra buen ajuste en algunos municipios, el rendimiento global sigue siendo inestable. El promedio del R ² negativo refleja que la regresión no supera la predicción con la media. El modelo tampoco generaliza bien en municipios con alta variabilidad. Se recomienda ajustar hiperparámetros o explorar modelos más robustos como Random Forest o XGBoost.
Random Forest	MAE: 19,84 / RMSE: 28,20 / R ² : -0,4042 / MAPE: 5,12	Precisión mixta. El modelo mejora el MAE y MAPE promedio frente a modelos anteriores y logra un buen ajuste en municipios como Villavicencio. Sin embargo, la regresión sigue siendo inestable en otras regiones. Se recomienda ajustar hiperparámetros e incorporar más variables para mejorar la generalización.
Gradient Boosting (XGBoost)	MAE: 26,88 / RMSE: 35,55 / R ² : -1,2637 / MAPE: 6,77	Desempeño inconsistente. Buenos resultados puntuales (ej. Villavicencio), pero el promedio del R ² sigue siendo negativo. El modelo no generaliza bien en municipios con alta variabilidad. Se recomienda ajuste de hiperparámetros o explorar modelos más robustos como redes neuronales.

Support Vector Regression (SVR)	MAE: 21,93 / RMSE: 34,46 / R^2 : -1,3338 / MAPE: 5,43	Desempeño irregular. Aunque el modelo logra un buen ajuste en municipios como Villavicencio, Bogotá y Cali, el rendimiento global sigue siendo inestable. El R^2 promedio es negativo, indicando que en la mayoría de municipios la regresión no supera la predicción con la media. Existen casos con errores extremos como Bucaramanga, Pereira y Bogotá D.C. Se recomienda ajustar hiperparámetros o usar modelos no lineales más robustos.
MLP (Perceptrón Multicapa)	MAE: 9,94 / RMSE: 14,74 / R^2 : -1,60 / MAPE: 2,50	Buen desempeño en la mayoría de municipios, con errores bajos y R^2 altos. Sin embargo, la regresión colapsa en Bogotá D.C. (R^2 : -21,46) y afecta fuertemente el promedio general. Se recomienda revisar outliers o reentrenar con mayor regularización para evitar sobreajuste en regiones inestables.
LSTM (Long Short-Term Memory)	MAE: 6,78 / RMSE: 8,99 / R^2 : 0,54 / MAPE: 1,76	Buen desempeño general del modelo LSTM en la mayoría de municipios, con errores bajos y valores R^2 superiores a 0.9. Sin embargo, al igual que en modelos anteriores, el rendimiento en Bogotá D.C. fue inestable y presenta un valor R^2 negativo, lo que indica una regresión peor que la media. Se recomienda revisar este municipio por posibles outliers o reentrenar con mayor regularización. El promedio general muestra un R^2 de 0.54, mejorando notablemente respecto a modelos como MLP, SVR o Random Forest.
GRU (Gated Recurrent Unit)	MAE: 8,36 / RMSE: 10,95 / R^2 : 0,19 / MAPE: 2,16	Buen rendimiento general en municipios como Medellín, Cali y Villavicencio, con valores R^2 mayores a 0,9. El modelo logra capturar bien la dinámica temporal del precio, y reduce el MAE y el MAPE respecto a modelos como MLP o Random Forest. Sin embargo, presenta inestabilidad en Bogotá D.C. (R^2 = -5,39) que afecta significativamente el R^2 promedio. Se recomienda revisar la calidad de los datos o ajustar la regularización. El rendimiento promedio es competitivo y representa una mejora respecto a otros modelos no recurrentes.

Tabla 7. Resultados Modelos Predictivos.

Capítulo 5. DISCUSIÓN

Durante el desarrollo del proyecto se evaluaron múltiples enfoques para la estimación de precios de electricidad en el mercado no regulado colombiano, con el objetivo de identificar un modelo preciso, robusto y generalizable que se adapte a la dinámica compleja de este mercado. A lo largo del proceso se abordaron desafíos técnicos, se validaron modelos predictivos y se tomaron decisiones metodológicas clave que vale la pena discutir.

➤ **Adaptación y cambios durante el proyecto.**

Inicialmente, la metodología planteada se basaba en el uso de modelos estadísticos clásicos y de aprendizaje automático. Sin embargo, a medida que avanzó el proyecto, fue necesario incorporar técnicas más avanzadas como redes neuronales profundas (LSTM y GRU), debido a la necesidad de capturar la alta granularidad horaria de los datos y las complejas relaciones no lineales presentes en el comportamiento del mercado.

Además, el proceso de integración de datos representó un reto considerable. La consolidación del dataset unificado, exigió un esfuerzo significativo en la limpieza, normalización y armonización de formatos, ya que los archivos descargados desde la plataforma de XM, empresa especializada en la gestión de sistemas de tiempo real y la administración del mercado de energía mayorista en Colombia, presentaban estructuras variables a lo largo de los años. A esto se sumó la complejidad de integrar datos de temperatura media mensual por municipio y de IPC mensual, lo cual requirió procesos de estimación, imputación y validación cruzada para asegurar la consistencia temporal y espacial de las variables exógenas.

➤ **Limitaciones del estudio.**

La calidad y consistencia de los datos históricos fue una limitación recurrente. La necesidad de reconstruir formatos, inferir días faltantes o corregir valores extremos (como los encontrados en Bogotá D.C.), afectó la calidad de entrada de algunos modelos.

El proceso de obtención de datos climáticos fue parcial en algunos municipios, debido a la falta de cobertura meteorológica directa. En esos casos se utilizaron estimaciones departamentales.

➤ **Impacto del proyecto y consideraciones finales.**

A pesar de las dificultades enfrentadas, la metodología empleada demostró ser eficaz y adaptable. El uso de modelos avanzados como GRU permitió alcanzar un nivel de precisión competitivo, abriendo la puerta a aplicaciones reales en entornos de toma de decisiones. La experiencia adquirida en el tratamiento y estructuración de datos heterogéneos constituye otro aporte valioso del proyecto.

Capítulo 6. CONCLUSIONES

6.1 Conclusiones del trabajo

- El proyecto logró cumplir satisfactoriamente el objetivo general de desarrollar un modelo predictivo, robusto y preciso, para estimar los precios de electricidad en el mercado no regulado colombiano. A través de un proceso iterativo de exploración, modelado y validación, se implementaron y evaluaron múltiples modelos estadísticos, de Machine Learning y Deep Learning, destacando diferencias notables en su desempeño.
- Los modelos clásicos como ARIMA/SARIMA presentaron limitaciones significativas en la captura de dinámicas complejas del mercado, reflejadas en bajos valores de R^2 y errores elevados. Por su parte, modelos como Árboles de decisión, SVR y XGBoost ofrecieron algunas mejoras puntuales, pero se mostraron inestables o poco generalizables entre municipios.
- El desempeño más destacado fue obtenido con modelos de redes neuronales profundas. En particular, LSTM alcanzó un R^2 promedio de 0.54, con resultados altamente precisos en municipios como Villavicencio, Medellín y Cali. El modelo GRU, aunque ligeramente inferior en métricas globales, también mostró un comportamiento sólido y estable en contextos con alta variabilidad temporal.
- Se evidenció, sin embargo, que municipios con alta variación o datos extremos (como Bogotá D.C.), requieren enfoques más robustos de regularización o detección de outliers, ya que algunos modelos presentaron colapsos significativos en esos contextos.
- En conjunto, los resultados demuestran que los modelos basados en redes neuronales recurrentes (RNN) como LSTM y GRU, ofrecen el mejor equilibrio entre precisión y estabilidad, representando una alternativa viable y poderosa para la estimación de precios en entornos no regulados.

6.2 Conclusiones personales

- La ejecución de este proyecto representó una experiencia académica, técnica y personal profundamente enriquecedora. Desde el inicio, la naturaleza del problema —la estimación de precios en un mercado energético volátil y no regulado— me exigió desarrollar una comprensión más profunda de los sistemas eléctricos colombianos, así como de las herramientas avanzadas de análisis de datos y modelado predictivo.
- Uno de los mayores desafíos fue la recolección, limpieza, estandarización y unificación de los datos históricos. La heterogeneidad de formatos descargados desde la plataforma XM empresa especializada en la gestión de sistemas de tiempo real y la administración del mercado de energía mayorista en Colombia, así como la integración de variables exógenas como temperatura e IPC, implicaron un trabajo detallado y constante de transformación y validación, que fortaleció mis competencias en ingeniería de datos.

- A nivel técnico, fue muy valioso comparar diferentes paradigmas de modelado desde lo clásico (como ARIMA), pasando por algoritmos de Machine Learning, hasta llegar a arquitecturas de Deep Learning como LSTM y GRU. Esta comparación me permitió entender con mayor claridad las ventajas y limitaciones de cada enfoque y tomar decisiones fundamentadas sobre cuál modelo aplicar en cada contexto.
- Finalmente, este proyecto no solo me permitió aplicar conocimientos adquiridos a lo largo del máster, sino que también me dio una visión más clara del potencial que tiene la inteligencia artificial para resolver problemas reales del sector energético en Colombia. Considero que este trabajo puede ser la base para desarrollos futuros con impacto técnico y estratégico, y ha reforzado mi motivación por seguir profundizando en el campo del análisis de datos.

Capítulo 7. FUTURAS LÍNEAS DE TRABAJO

A partir del desarrollo del presente proyecto y de los resultados obtenidos, se identifican múltiples oportunidades para profundizar o ampliar el trabajo realizado. Estas futuras líneas de investigación permitirían mejorar la precisión de los modelos, ampliar su aplicabilidad y facilitar su adopción en entornos reales del sector energético colombiano. Entre las más relevantes se destacan:

➤ **Mejora en la calidad y consistencia de los datos históricos.**

La obtención y unificación del dataset implicó un tratamiento exhaustivo de formatos heterogéneos provenientes de la plataforma XM empresa especializada en la gestión de sistemas de tiempo real y la administración del mercado de energía mayorista en Colombia, así como la integración de variables exógenas como temperatura e IPC. Futuras investigaciones podrían automatizar este proceso de consolidación y validar con fuentes oficiales adicionales para reducir posibles errores u omisiones.

➤ **Profundización en el análisis por municipio.**

Algunos municipios, como Bogotá D.C., presentaron resultados inestables debido a alta variabilidad, outliers o ruido en las series. Es recomendable desarrollar modelos personalizados o segmentar los municipios por características comunes para mejorar la precisión local.

➤ **Incorporación de nuevas variables exógenas.**

Aunque se incluyeron variables como temperatura e IPC, podrían añadirse otras fuentes de información como niveles de precipitación, viento, precios internacionales de gas y carbón, nivel de reservas hídricas, generación renovable o indicadores económicos del sector eléctrico. Esto podría aumentar la capacidad predictiva en horizontes más largos.

➤ **Exploración de modelos híbridos y arquitecturas profundas.**

Se evidenció que modelos como LSTM y GRU superaron significativamente a modelos clásicos (ARIMA, árboles, SVR), especialmente en municipios con buena calidad de datos. Es recomendable probar arquitecturas más complejas como Encoder-Decoder, Transformers o modelos híbridos que combinen estadísticas clásicas con Deep Learning.

➤ **Desarrollo de una plataforma de visualización y monitoreo.**

Con el modelo ya entrenado, una futura línea sería la creación de dashboards interactivos o sistemas de alerta temprana para apoyar decisiones en tiempo real por parte de comercializadores o grandes consumidores de energía.

➤ **Validación en entorno real con datos en línea.**

Finalmente, sería valioso implementar el modelo en un entorno semiproducción que reciba datos actualizados en tiempo real (vía API o scrapping), con el fin de probar su robustez, estabilidad y capacidad de adaptación frente a nuevas condiciones del mercado.

Capítulo 8. REFERENCIAS

AENOR. (2010). *AEN/CTN 157 - PROYECTOS*. Recuperado el 25 de abril de 2013, de Normas y Publicaciones:

<http://www.aenor.es/aenor/normas/ctn/fichactn.asp?codigonorm=AEN/CTN%20157>

Barrientos, J., Rodas, E., Velilla, E., Lopera, M., & Villada, F. (2012). Modelo para el pronóstico del precio de la energía eléctrica en Colombia. *Lecturas de Economía*, 77, 91–127.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

DANE. (2018). *IPC - Índice de Precios al Consumidor*. Obtenido de [https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/ipc-](https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/ipc-historico?highlight=WyJkYXRlliwiZGF0b3MiLCJkYXRlliwiZGF0byIsImRhZGFuliwiaGlzdFx1MdBmM3JpY2EiLCJoaXN0XHUwMGYzcmJlb3MiLCJoaXN0XHUwMGYzcmJlYXMiLC)

[historico?highlight=WyJkYXRlliwiZGF0b3MiLCJkYXRlliwiZGF0byIsImRhZGFuliwiaGlzdFx1MdBmM3JpY2EiLCJoaXN0XHUwMGYzcmJlb3MiLCJoaXN0XHUwMGYzcmJlYXMiLC](https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/ipc-historico?highlight=WyJkYXRlliwiZGF0b3MiLCJkYXRlliwiZGF0byIsImRhZGFuliwiaGlzdFx1MdBmM3JpY2EiLCJoaXN0XHUwMGYzcmJlb3MiLCJoaXN0XHUwMGYzcmJlYXMiLC)

Gallón, S., & Barrientos, J. (2021). Forecasting the Colombian electricity spot price under a functional approach. *International Journal of Energy Economics and Policy*, 11(2), 67–74.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Hurtado, L., Quintero, O., & García, J. (2014). Estimación del precio de oferta de la energía eléctrica en Colombia mediante inteligencia artificial. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 18, 54-87.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

IDEAM. (2010). *Datos Abiertos Clima Colombia*. Obtenido de <https://visualizador.ideam.gov.co/CatalogoObjetos/geo-open-data/search?search=&theme=SEGUIMIENTO%20Y%20CARACTERIZACI%C3%93N%20DEL%20CLIMA&group=>

Miró Julià, J. (2010). *Recursos para aprender a escribir*. Recuperado el septiembre de 2012, de <http://bioinfo.uib.es/~joemiro/RecEscr/manual.pdf>

UNE 157001. (2002). *Criterios generales para la elaboración de proyectos*. Recuperado el 25 de abril de 2013, de Escuela Universitaria de Ingeniería de Vitoria:

http://www.coiib.es/coiib/documentos/DocumentosContenidos/Gu%C3%ADa%20de%20elaboraci%C3%B3n%20de%20proyectos/2-Electricidad/5_PNE_157701_Criterios.pdf

Velásquez, J. D., Franco, C. J., & García, H. A. (2009). Un modelo no lineal para la predicción de la demanda mensual de electricidad en Colombia. *Ingeniería e Investigación*, 29(2), 95-100.

Velásquez, J. D., Franco, C. J., & Olaya, Y. (2010). Predicción de los precios promedios mensuales de contratos despachados en el mercado mayorista de electricidad en Colombia. *Revista EIA*, 12, 101–113.

XM. (2021). *Precio promedio y energía transada*. Obtenido de <https://www.xm.com.co/transacciones/registros/registro-contratos/informe-especial/precio-promedio-y-energia-transada>

Zhang, G., Eddy Patuwo, B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62.

Capítulo 9. ANEXOS

➤ **Anexo 1. Dataset.ipynb**

Cuaderno de Jupyter utilizado para la exploración, depuración y validación del dataset base. Incluye el código para importar, visualizar y verificar la consistencia de los datos históricos extraídos del mercado eléctrico colombiano.

➤ **Anexo 2. dataset_unificado_2006_2021.csv**

Archivo CSV con la base de datos consolidada entre 2006 y 2021. Contiene información histórica diaria sobre precios promedio, demanda, agente comercializador, ubicación geográfica y otros atributos energéticos relevantes.

➤ **Anexo 3. dataset_unificado_2006_2021_con_temperatura_final_con_ipc.xlsx**

Archivo Excel que extiende el dataset anterior incorporando dos variables exógenas clave: temperatura media mensual por municipio e IPC mensual, alineadas por mes y año.

➤ **Anexo 4. Modelos.ipynb**

Cuaderno de Jupyter con la implementación de los modelos predictivos candidatos (SARIMA, Random Forest, XGBoost, SVM, LSTM, GRU, entre otros). Incluye procesamiento de datos, entrenamiento de modelos, evaluación mediante métricas MAE, RMSE, MAPE y R^2 , y visualización de resultados.

[PÁGINA INTENCIONADAMENTE EN BLANCO]