



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO

MAESTRÍA EN CIENCIAS MENCIÓN INFORMÁTICA

TESIS

**MODELO DE PREDICCIÓN DE LA DEMANDA DE ENERGÍA
ELÉCTRICA EN EL PERÚ UTILIZANDO REDES
NEURONALES LSTM**

**PARA OPTAR AL GRADO ACADÉMICO DE MAESTRO EN
CIENCIAS MENCIÓN INFORMÁTICA**

AUTOR:

Br. VANESA LAVILLA ALVAREZ

ASESOR:

Dr. RONY VILLAFUERTE SERNA

CODIGO ORCID:

0000-0003-4607-522X

CUSCO – PERÚ

2025

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, Asesor del trabajo de investigación/tesis titulada: MODELO DE PREDICCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA EN EL PERÚ UTILIZANDO REDES NEURONALES LSTM

presentado por: VANESA LAVILLA ALVAREZ con DNI Nro. 44731982
para optar el título profesional/grado académico de MAESTRO EN CIENCIAS MENCION INFORMÁTICA

Informo que el trabajo de investigación ha sido sometido a revisión por 02 veces, mediante el Software Antiplagio, conforme al Art. 6° del **Reglamento para Uso de Sistema Antiplagio de la UNSAAC** y de la evaluación de originalidad se tiene un porcentaje de 1 %.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis.

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y adjunto la primera página del reporte del Sistema Antiplagio.

Cusco, 12 de AGOSTO de 2025



Firma

Post firma RONY VILLAFUERTE SERNA

Nro. de DNI 23957778

ORCID del Asesor 0000 - 0003 - 4607 - 522X

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: oid: 27259:482155640

Vanesa Lavilla Alvarez

MODELO DE PREDICCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA EN EL PERÚ UTILIZANDO REDES NEURONALES L...

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:482155640

Fecha de entrega

12 ago 2025, 9:44 a.m. GMT-5

Fecha de descarga

12 ago 2025, 9:49 a.m. GMT-5

Nombre de archivo

MODELO DE PREDICCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA EN EL PERÚ UTILIZANDO RED....pdf

Tamaño de archivo

6.8 MB

124 Páginas

22.163 Palabras

125.180 Caracteres

1% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Coincidencias menores (menos de 20 palabras)

Exclusiones

- ▶ N.º de coincidencias excluidas

Fuentes principales

- 1%  Fuentes de Internet
- 0%  Publicaciones
- 1%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO

INFORME DE LEVANTAMIENTO DE OBSERVACIONES A TESIS

Dra. NELLY AYDE CAVERO TORRE, Directora (e) General de la Escuela de Posgrado, nos dirigimos a usted en condición de integrantes del jurado evaluador de la tesis intitulada MODELO DE PREDICCIÓN DE LA DEMANDA DE ENERGÍA ELÉCTRICA EN EL PERÚ UTILIZANDO REDES NEURONALES LSTM de la Br. Br. VANESA LAVILLA ALVAREZ. Hacemos de su conocimiento que el (la) sustentante ha cumplido con el levantamiento de las observaciones realizadas por el Jurado el día DOS DE JULIO DE 2025.

Es todo cuanto informamos a usted fin de que se prosiga con los trámites para el otorgamiento del grado académico de MAESTRO EN CIENCIAS MENCIÓN INFORMÁTICA.

Cusco, 08 de agosto del 2025



DR. LUIS BELTRAN PALMA TTITO
Primer Replicante



DR. JAVIER ARTURO ROZAS HUACHO
Segundo Replicante



DR. CARLOS FERNANDO MONTOYA CUBAS
Primer Dictaminante



MGT. RAY DUEÑAS JIMENEZ
Segundo Dictaminante

Dedicatoria

*Esta tesis va dedicada con mucho amor para mi familia Richarth y Hana Akira
Este paso es por ustedes.*

Vanesa.

Agradecimientos

Agradezco sinceramente a mi familia por su apoyo incondicional y aliento constante a lo largo de este proceso. A mis profesores, quienes con su guía y conocimiento me ayudaron a enriquecer este trabajo. A mis amigos y compañeros, por su motivación y compañía en cada etapa.

Gracias a todos por hacer posible la realización de esta tesis.

Vanesa.

Índice general

Dedicatoria	II
Agradecimientos	III
Resumen	XII
Abstract	XIII
Introducción	1
1. Planteamiento del problema	3
1.1. Situación problemática	3
1.2. Formulación del problema	5
1.2.1. Problema general	5
1.2.2. Problemas específicos	5
1.3. Justificación de la investigación	5
1.4. Objetivos de la investigación	7
1.4.1. Objetivo general	7
1.4.2. Objetivos específicos	7
2. Marco Teórico Conceptual	8
2.1. Bases teóricas	8
2.1.1. Inteligencia artificial	8
2.1.2. Machine learning	8
2.1.3. Redes neuronales	9

2.2.	Marco conceptual	9
2.2.1.	Series temporales	9
2.2.2.	Anomalías en series temporales	10
2.2.3.	Demanda de energía eléctrica	11
2.2.4.	Pandemia COVID-19 en Perú	12
2.2.5.	Red neuronal LSTM	14
2.3.	Antecedentes Empíricos de la Investigación	16
2.3.1.	Dinh et al. (2024), <i>Electrical Energy Demand Forecasting using Time Series in LSTM and CNN-LSTM Models in Deep Learning Applications</i> , Electrical Energy Demand Forecasting using Time Series in LSTM and CNN-LSTM, USA.	16
2.3.2.	Rahman et al. (2024), <i>Time Series-Based Analysis of Energy Consumption: Forecasting and Anomaly Detection Using LSTM and Isolation Forest</i> , International Journal of Engineering Research and Technology, USA.	17
2.3.3.	Hanif et al. (2021), <i>Power Consumption Predictive Analytics and Automatic Anomaly Detection Using CNN-LSTM</i> , 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE).	18
2.3.4.	Blázquez G. et al. (2021), <i>A review on outlier/anomaly detection in time series data</i> , ACM computing surveys (CSUR), USA.	19
2.3.5.	Ergen and Kozat (2019), <i>Unsupervised and Semi-supervised Anomaly Detection with LSTM Neural Networks</i> , IEEE transactions on neural networks and learning systems, USA.	19
2.3.6.	Niu et al. (2020), <i>LSTM-Based VAE-GAN for Time-Series Anomaly Detection</i> , Sensors 2020, China.	20
2.3.7.	Bakhtawar M. (2019), <i>Anomaly detection in electricity demand time series data</i> , Kth Royal Institute of Technology School of Electrical Engineering and Computer Science, Suecia.	21

2.3.8.	Malhotra et al. (2015), <i>Long Short Term Memory Networks for Anomaly Detection in Time Series</i> , ESSANN 2015 23rd European Symposium on Artificial Neural Networks Computational Intelligence and Machine Learning, Belgica.	22
2.3.9.	Pan et al. (2022), <i>High-Dimensional Energy Consumption Anomaly Detection: A Deep Learning - Based Method for Detecting Anomalies</i> , School of Mechanical and Automatic, Zhejiang Sci - Tech University, Hangzhou, China.	23
2.3.10.	Cui et al. (2023), <i>A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting</i> , Department of Computer Science, Changzhi University, Changzhi, Shanxi, China.	24
3.	Metodología	25
3.1.	Ámbito de estudio: Localización política y geográfica	25
3.2.	Tipo y nivel de investigación	25
3.3.	Metodología de la investigación	26
3.4.	Unidad de análisis	28
3.5.	Población de estudio	29
3.6.	Técnicas de recolección de información	29
3.7.	Métricas para evaluar el modelo de regresión	29
3.7.1.	Raíz del error cuadrático medio	29
3.7.2.	Error absoluto medio	30
3.7.3.	Coefficiente de determinación	30
3.8.	Herramientas informáticas utilizadas	32
4.	Resultados	33
4.1.	Recolección y preparación de datos	33
4.1.1.	Acceso y recolección de datos	33
4.1.2.	Descripción del archivo de datos	34
4.1.3.	Descripción general de los datos	35

4.1.4.	Exploración inicial	35
4.1.5.	Análisis de gráficos iniciales	36
4.2.	División temporal	39
4.2.1.	Definición de las etapas	39
4.2.2.	Análisis estadístico por etapa	40
4.2.3.	Análisis estadístico inferencial de la demanda de energía eléctrica	42
4.3.	Normalización de datos	47
4.4.	Definición y entrenamiento	49
4.4.1.	Definición del modelo LSTM seleccionado	49
4.4.2.	Creación de la base de datos supervisada	49
4.4.3.	Justificación del modelo seleccionado	52
4.4.4.	Configuración de hiperparámetros	52
4.4.5.	Entrenamiento del modelo	56
4.4.6.	Evaluación del modelo	57
4.4.7.	Interpretación de resultados del modelo de predicción	57
4.5.	Detección de anomalías	58
4.5.1.	Metodología aplicada	58
4.5.2.	Umbral para identificar anomalías	58
4.5.3.	Cantidad de anomalías detectadas	59
4.5.4.	Visualización de anomalías	61
4.6.	Análisis post pandemia	63
4.7.	Evaluación de resultados	71
4.8.	Análisis adicional	73
5.	Discusión de resultados	77
5.1.	Rendimiento del modelo predictivo	77
5.2.	Identificación de anomalías	78
5.3.	Implicancias para la planificación energética	78
5.4.	Limitaciones del estudio	78
5.5.	Propuestas futuras	79

5.6. Comparación con estudios del estado del arte	79
5.7. Comparación de resultados con el estado del arte	81
6. Conclusiones y recomendaciones	84
6.1. Conclusiones	84
6.2. Recomendaciones	86
Bibliografía	88
Anexos	93

Índice de Tablas

4.1. Estadísticas descriptivas de la variable EJECUTADO.	35
4.2. Resumen estadístico de la variable EJECUTADO en los tres periodos analizados.	41
4.3. Hiperparámetros seleccionados y sus justificaciones.	55
4.4. Métricas con datos durante y post-pandemia.	71
4.5. Métricas con datos de marzo a diciembre de 2024.	73
4.6. Comparación de métricas de este estudio con datos predecidos del COES.	73
5.1. Comparación de resultados con estudios del estado del arte	83

Índice de figuras

2.1. Componentes de una serie de tiempo.	10
2.2. Tipos de anomalías en series de tiempo.	11
2.3. Demanda de energía eléctrica en el Perú.	12
2.4. Coronavirus en el Perú.	13
2.5. Célula de LSTM.	15
3.1. Pasos metodología de investigación	27
3.2. Demanda de energía eléctrica en Perú: lapso de 2 semanas.	28
4.1. Archivo de demanda de energía eléctrica.	34
4.2. Demanda de energía eléctrica en Perú 2018 - 2024.	36
4.3. Descomposición de la serie temporal.	37
4.4. Histograma y diagrama de caja de la demanda de energía eléctrica. . .	38
4.5. Etapas de la serie temporal del estudio	40
4.6. Boxplot por etapa	41
4.7. Resultados prueba tukey HSD	46
4.8. Demanda eléctrica por fase	47
4.9. Normalización de la serie temporal	48
4.10. Creación base de datos supervisada	50
4.11. Muestra del conjunto de entrenamiento	51
4.12. Comportamiento demanda de energía eléctrica por día	53
4.13. Comportamiento demanda de energía eléctrica por semana	53
4.14. Comportamiento demanda de energía eléctrica por año	54
4.15. Anomalías durante la pandemia	60

4.16. Anomalías post-pandemia	60
4.17. Comparación de datos reales y predicciones durante la pandemia.	62
4.18. Comparación de datos reales y predicciones post-pandemia	64
4.19. Histograma de distribución de errores absolutos durante y post-pandemia	65
4.20. Distribución de anomalías por mes y promedio de amplitud por mes	67
4.21. Comparación de anomalías consecutivas etapas durante y post-pandemia	68
4.22. Anomalías por hora puntual y consecutivas en los periodos durante y post-pandemia	70
4.23. Comparación de valores reales, predichos por el modelo y valores pro- gramados por el COES	75

Resumen

La pandemia de COVID-19 generó alteraciones sin precedentes en los patrones de consumo de energía eléctrica en el Perú, debido a las medidas de confinamiento y teletrabajo. Estos cambios plantearon la necesidad de identificar anomalías prolongadas en la demanda de energía eléctrica durante y después del periodo crítico de la pandemia. Esta investigación tuvo como objetivo desarrollar un modelo predictivo basado en redes neuronales LSTM para estimar la demanda de energía eléctrica en el Perú, evaluando su desempeño y su utilidad en la identificación de anomalías. Se diseñó e implementó un modelo LSTM utilizando datos históricos de demanda eléctrica registrados cada 30 minutos entre los años 2018 y 2024. Los datos se dividieron en tres etapas: pre-pandemia, durante la pandemia y post-pandemia. El modelo fue entrenado con datos del periodo pre-pandemia y evaluado mediante métricas como RMSE, MAE y R^2 . Posteriormente, se emplearon los errores de predicción para identificar anomalías en los periodos durante y post-pandemia, se usó el umbral percentil 95. Los resultados mostraron un RMSE inferior al 2% de la demanda media y un R^2 superior al 97%, evidenciando una alta capacidad predictiva del modelo. Se detectaron 1,733 anomalías durante la pandemia, y 1,916 en la etapa post-pandemia. El mayor número de anomalías post-pandemia se vinculó a la inestabilidad política y social que afectaron la actividad económica y consumo eléctrico, aunque con menor amplitud. Este estudio confirma la eficacia de las redes LSTM para detectar anomalías y anticipar comportamientos atípicos en contextos de crisis.

PALABRAS CLAVE: demanda de energía eléctrica, anomalías, redes neuronales LSTM, COVID-19, series temporales.

Abstract

The COVID-19 pandemic caused unprecedented changes in electricity consumption patterns in Peru, due to lockdown measures and the widespread adoption of remote work. These disruptions highlighted the need to identify prolonged anomalies in electricity demand during and after the critical period of the pandemic. This research aimed to develop a predictive model based on Long Short-Term Memory (LSTM) neural networks to estimate electricity demand in Peru, evaluating its performance and usefulness in anomaly detection. An LSTM model was designed and implemented using historical electricity demand data recorded every 30 minutes between 2018 and 2024. The dataset was divided into three stages: pre-pandemic, during the pandemic, and post-pandemic. The model was trained using pre-pandemic data and evaluated with metrics such as RMSE, MAE, and R^2 . Subsequently, prediction errors were used to detect anomalies in the during- and post-pandemic stages, applying the 95th-percentile threshold. The results showed an RMSE below 2% of the average demand and an R^2 greater than 97%, evidencing the model's high predictive capability. A total of 1,733 anomalies were detected during the pandemic and 1,916 in the post-pandemic stage. The higher number of post-pandemic anomalies was linked to political and social instability that affected economic activity and electricity consumption, albeit with lower amplitude. This study confirms the effectiveness of LSTM networks for detecting anomalies and anticipating atypical behaviors in electricity demand under health or economic crisis contexts.

Keywords: electricity demand, anomalies, LSTM neural networks, COVID-19, time series.

Introducción

Esta tesis, titulada "Modelo para la detección de anomalías en la demanda de energía eléctrica en el Perú utilizando redes neuronales LSTM en los periodos pre-pandemia, durante la pandemia y post-pandemia de COVID-19", tiene como objetivo desarrollar un modelo predictivo para estimar la demanda de energía eléctrica y analizando con este los cambios en el comportamiento de la demanda energética a través de la detección de anomalías, empleando redes neuronales de memoria a corto y largo plazo (LSTM). La investigación busca identificar si los patrones anómalos provocados por la pandemia de COVID-19 persisten en la etapa post-pandemia, afectando la estabilidad y gestión del sistema eléctrico peruano. En el Capítulo 1, se presenta el planteamiento del problema, los objetivos generales y específicos, y la justificación del estudio, resaltando la necesidad de utilizar modelos de aprendizaje profundo para detectar eventos atípicos en la serie de demanda eléctrica. El Capítulo 2 desarrolla el marco conceptual, incluyendo fundamentos sobre inteligencia artificial, aprendizaje automático, redes neuronales, series temporales y detección de anomalías, además de una revisión del contexto energético peruano y del estado del arte nacional e internacional en predicción energética y análisis de anomalías. La metodología, detallada en el Capítulo 3, enmarca la investigación como de tipo aplicada, con un nivel predictivo, enfoque cuantitativo y diseño experimental supervisado. Se utilizaron 105,216 registros horarios de demanda eléctrica del COES, comprendidos entre marzo de 2018 y marzo de 2024, los cuales fueron preprocesados, normalizados y divididos en tres periodos (pre-pandemia, durante la pandemia y post-pandemia). El modelo LSTM fue

entrenado con datos del periodo pre-pandemia y aplicado a los otros dos, utilizando métricas como RMSE, MAE y R^2 para evaluar su desempeño y detectar anomalías a partir de umbrales derivados de los errores de predicción. El Capítulo 4 presenta los resultados obtenidos. Se identificaron 1,733 anomalías durante la pandemia, asociadas principalmente a los confinamientos y la interrupción de actividades económicas, mientras que en el periodo post-pandemia se detectaron 1,916 anomalías, lo cual indica la persistencia de eventos irregulares, aunque con menor intensidad. El modelo alcanzó un RMSE inferior al 2% de la media y un R^2 superior al 97%, lo que demuestra su alta capacidad predictiva. Se incluyeron visualizaciones comparativas que muestran la precisión del modelo frente a enfoques tradicionales. El Capítulo 5, desarrolla la discusión de los resultados. En él se contrastan los hallazgos de esta tesis con estudios previos del estado del arte, se analizan las posibles causas contextuales de los patrones anómalos —como la recuperación económica, crisis políticas internas y variaciones estacionales, y se reflexiona sobre las implicancias del comportamiento post-pandemia en la planificación energética nacional. Además, se incluyen análisis mensuales y una tabla comparativa que fortalecen la interpretación de las anomalías detectadas. Finalmente, en el Capítulo 6, se presentan las conclusiones y recomendaciones. Se confirma que el modelo propuesto no solo es eficaz para detectar patrones atípicos, sino también útil para anticipar eventos futuros en la gestión del sistema eléctrico. Se recomienda incorporar variables externas (como temperatura, indicadores económicos o días festivos), aplicar el modelo en regiones específicas del país y explorar arquitecturas híbridas como CNN-LSTM o LSTM con mecanismos de atención. Este estudio establece un marco metodológico robusto, replicable y adaptable, que puede ser de gran utilidad para las entidades del sector energético frente a escenarios disruptivos futuros.

Capítulo 1

Planteamiento del problema

1.1. Situación problemática

La pandemia de la COVID-19, que se extendió desde 2020 hasta mediados de 2022, tuvo un impacto significativo en las actividades económicas y sociales a nivel mundial. En el Perú, las medidas implementadas como el confinamiento obligatorio, las restricciones de movilidad y los cambios en los patrones de trabajo y vida cotidiana modificaron, de manera directa e indirecta, los hábitos de consumo energético de la población.

La demanda de energía eléctrica en el Perú, al igual que en otros países, está influenciada por fluctuaciones diarias, estacionales y factores económicos. Su análisis representa un desafío técnico debido a la necesidad de garantizar el suministro continuo y eficiente dentro del Sistema Eléctrico Interconectado Nacional (SEIN), gestionado por el Comité de Operación Económica del Sistema (COES). Para lograrlo, es esencial contar con herramientas precisas que permitan anticipar la demanda y detectar posibles anomalías que puedan comprometer la estabilidad del sistema.

En este contexto, la pandemia representa un evento disruptivo que pudo alterar los patrones normales de consumo eléctrico, no solo durante su periodo crítico, sino

también en el tiempo posterior. Identificar si estos cambios han generado anomalías persistentes en la serie temporal de la demanda de energía eléctrica es una necesidad relevante para la planificación energética a largo plazo.

Dado que la demanda eléctrica fue registrada cada 30 minutos, se dispone de una gran cantidad de datos. Sin embargo, esta riqueza informativa también presenta desafíos técnicos, especialmente en la diferenciación entre fluctuaciones normales (como las cíclicas o estacionales) y anomalías reales provocadas por eventos externos. Detectar anomalías prolongadas en contextos como este requiere métodos avanzados de análisis temporal.

En los últimos años, el aprendizaje profundo ha mostrado resultados prometedores en la detección de anomalías en distintos dominios como redes, IoT y salud (Pang et al., 2021; Al-amri et al., 2021; Fernando et al., 2020). No obstante, existe una carencia de estudios aplicados a series temporales prolongadas en variables estructurales como la demanda energética, especialmente en contextos donde los cambios no son abruptos, sino progresivos o indirectos.

Las redes neuronales LSTM (*Long Short-Term Memory*) se han consolidado como herramientas eficaces para modelar secuencias temporales complejas, dada su capacidad para capturar dependencias a largo plazo (Fischer and Krauss, 2018; Kratzert et al., 2022; Cissoko, 2024). En este estudio, se propone desarrollar un modelo predictivo basado en LSTM que permita estimar la demanda eléctrica en el Perú, evaluar su desempeño mediante métricas de regresión, y analizar las diferencias entre las predicciones y los valores reales con el fin de identificar anomalías que hayan surgido durante y después de la pandemia de COVID-19.

Comprender estas posibles alteraciones resulta crucial no solo para el diagnóstico del impacto de la pandemia, sino también para fortalecer la capacidad de respuesta ante futuros eventos disruptivos que puedan afectar el consumo energético nacional.

1.2. Formulación del problema

1.2.1. Problema general

¿Cómo desarrollar un modelo predictivo basado en redes neuronales LSTM que permita estimar con precisión la demanda de energía eléctrica en el Perú y que sea útil para identificar anomalías durante y después de la pandemia de COVID-19?

1.2.2. Problemas específicos

1. ¿Cómo diseñar e implementar un modelo predictivo basado en redes neuronales LSTM utilizando datos históricos de la demanda eléctrica en el Perú?
2. ¿Qué nivel de precisión alcanza el modelo predictivo desarrollado al ser evaluado mediante métricas de regresión?
3. ¿Qué anomalías pueden identificarse en la demanda de energía eléctrica al comparar las predicciones del modelo con los datos reales durante y después de la pandemia de COVID-19?

1.3. Justificación de la investigación

La pandemia de la COVID-19 ha representado un evento global sin precedentes, cuyas consecuencias se extendieron más allá del ámbito sanitario, impactando profundamente la economía, el comportamiento social y el uso de recursos esenciales como la energía eléctrica. En el Perú, los cambios en los patrones de consumo energético durante y después de la pandemia han planteado nuevos desafíos para la gestión y planificación del sistema eléctrico nacional.

La energía eléctrica es un insumo fundamental para el funcionamiento continuo

de hospitales, industrias, hogares y servicios públicos. Su disponibilidad confiable depende de una planificación adecuada basada en predicciones precisas de la demanda. En este contexto, desarrollar modelos predictivos robustos es una necesidad prioritaria, especialmente ante escenarios disruptivos como el vivido durante la pandemia.

A pesar de la importancia de este tema, existe una brecha de conocimiento respecto al impacto indirecto y prolongado de la pandemia sobre la demanda eléctrica, así como sobre la capacidad de los modelos actuales para anticipar estos cambios. Esta investigación busca contribuir al cierre de esa brecha mediante el desarrollo de un modelo predictivo basado en redes neuronales LSTM, una técnica avanzada de aprendizaje profundo que ha demostrado alta efectividad en el análisis de series temporales complejas (Zhao et al., 2021).

El modelo propuesto no solo permitirá estimar con precisión la demanda de energía eléctrica en el Perú, sino también identificar posibles anomalías comparando las predicciones con los valores reales durante y después de la pandemia. Este enfoque es fundamental para comprender mejor la dinámica energética post-pandemia y sentar las bases para una infraestructura energética más resiliente.

Finalmente, los resultados de este estudio aportarán evidencia valiosa para que entidades como el Comité de Operación Económica del Sistema (COES), empresas del sector eléctrico y formuladores de políticas puedan mejorar sus mecanismos de monitoreo, predicción y respuesta frente a situaciones críticas, fortaleciendo así la seguridad y sostenibilidad del sistema eléctrico nacional.

1.4. Objetivos de la investigación

1.4.1. Objetivo general

Desarrollar un modelo predictivo basado en redes neuronales LSTM para estimar la demanda de energía eléctrica en el Perú, evaluando su desempeño y utilidad en la identificación de anomalías durante y después de la pandemia de COVID-19.

1.4.2. Objetivos específicos

1. Diseñar e implementar un modelo predictivo basado en redes neuronales LSTM, utilizando datos históricos de la demanda eléctrica en Perú.
2. Evaluar la precisión del modelo predictivo mediante métricas de regresión.
3. Identificar y analizar anomalías comparando los valores predichos por el modelo con los datos reales durante y después de la pandemia.

Capítulo 2

Marco Teórico Conceptual

2.1. Bases teóricas

2.1.1. Inteligencia artificial

Russell and Norvig (2016), nos menciona que la inteligencia artificial es un campo de estudio de la informática que se ocupa del desarrollo de sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, la toma de decisiones, la comprensión del lenguaje y el aprendizaje.

2.1.2. Machine learning

Murphy (2012), indica que el aprendizaje automático es una rama de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos, identificar patrones y tomar decisiones con una intervención humana mínima .

2.1.3. Redes neuronales

Goodfellow et al. (2016), sobre las redes neuronales artificiales menciona que son modelos computacionales inspirados en la estructura del cerebro humano, compuestas por nodos (neuronas) organizados en capas, que se utilizan para procesar datos no lineales y complejos mediante conexiones ponderadas y funciones de activación.

2.2. Marco conceptual

2.2.1. Series temporales

Una serie temporal es una secuencia de n observaciones ordenadas y equidistantes cronológicamente sobre una característica llamada univariante o sobre varias características llamadas multivariantes de una unidad observable en diferentes momentos. (Box et al., 2016)

Representación matemática:

$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t : t = 1, \dots, N)$, donde y_t es la observación t ($1 \leq t \leq N$) de la serie.

Componentes

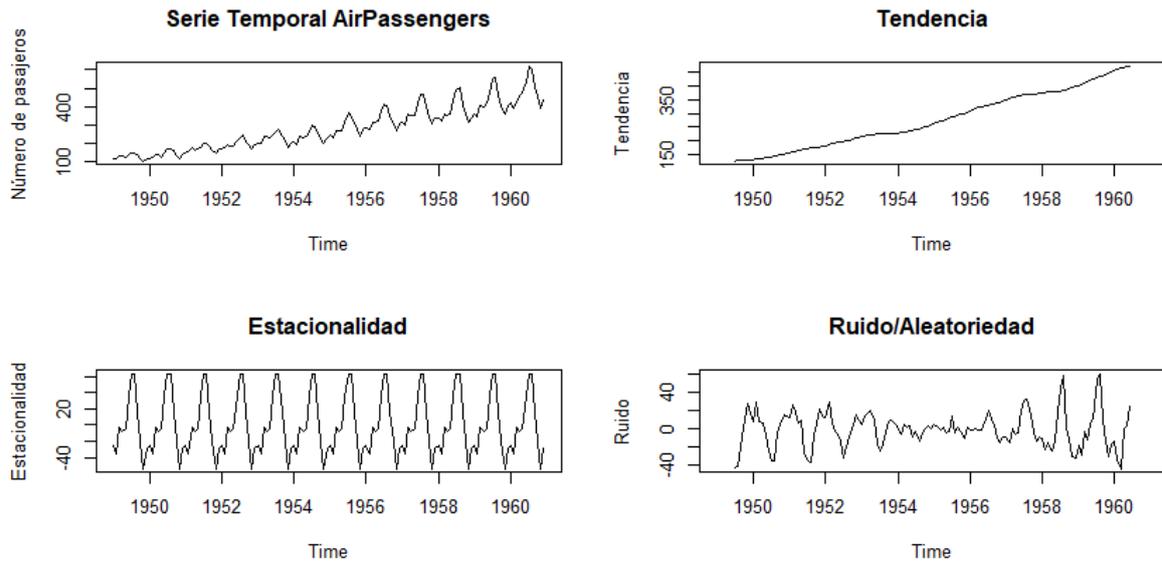
Los componentes de una serie de tiempo son (Chatfield and Xing, 2019):

- Tendencia, es un cambio a lo largo del tiempo.
- Estacionalidad, patrón estacional, que es un cambio regular y predecible en periodos fijos.
- Cíclico: oscilaciones que no tienen periodo fijo.
- Ruido: variaciones muy pequeñas que no pertenecen a la señal original

La Figura 2.1 ilustra un ejemplo de la serie temporal del número de pasajeros en vuelos. Se observa inicialmente la serie original, seguida de su descomposición en los componentes de tendencia, estacionalidad anual y residuos (ruido aleatorio).

Figura 2.1

Componentes de una serie de tiempo.



Fuente: Elaboración propia con RSTUDIO.

2.2.2. Anomalías en series temporales

Una anomalía en una serie de tiempo es un punto de datos o un conjunto de puntos que se desvían significativamente del patrón normal o esperado de la serie temporal. Los valores atípicos o excepcionales también son palabras para anomalías (Wu, 2016).

Tipos de anomalías (Teng et al., 2017):

- **Anomalías puntuales:** La más básica, es cuando el detector analiza cada punto y encuentra una anomalía, una observación notablemente diferente o cuando un punto sale del umbral normal.
- **Anomalías contextuales:** No solo examina la anomalía en sí, sino también el

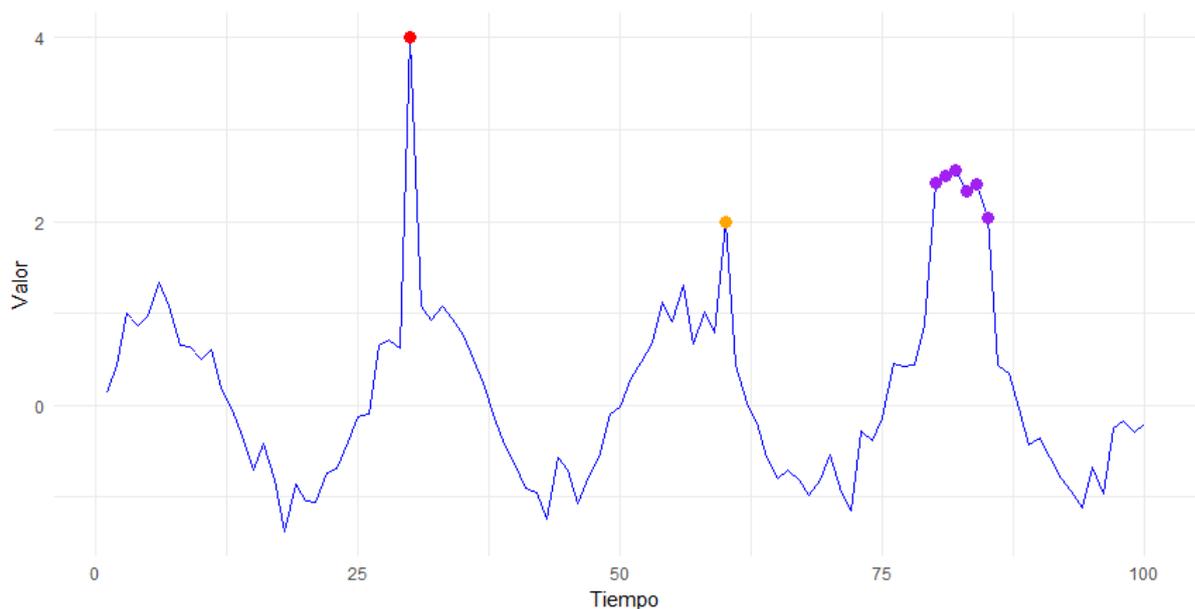
contexto en el que se presentó en otras series.

- **Anomalías colectivas:** Es la secuencia o conjunto de datos etiquetado como anomalías.

Por ejemplo en la Figura 2.2, se aprecia una serie con un patrón sinusoidal con algo de ruido para simular datos reales, el punto rojo, naranja y morados representan las anomalías puntuales, contextuales y colectivas respectivamente.

Figura 2.2

Tipos de anomalías en series de tiempo.



Fuente: Elaboración propia con RSTUDIO.

2.2.3. Demanda de energía eléctrica

El término "demanda de energía eléctrica" se refiere a la cantidad de energía eléctrica que necesitan los consumidores y las industrias en un momento y lugar específicos (Alonso, 2012). Es la cantidad total de energía eléctrica necesaria para satisfacer las necesidades de los usuarios en un área específica y durante un período de tiempo específico.

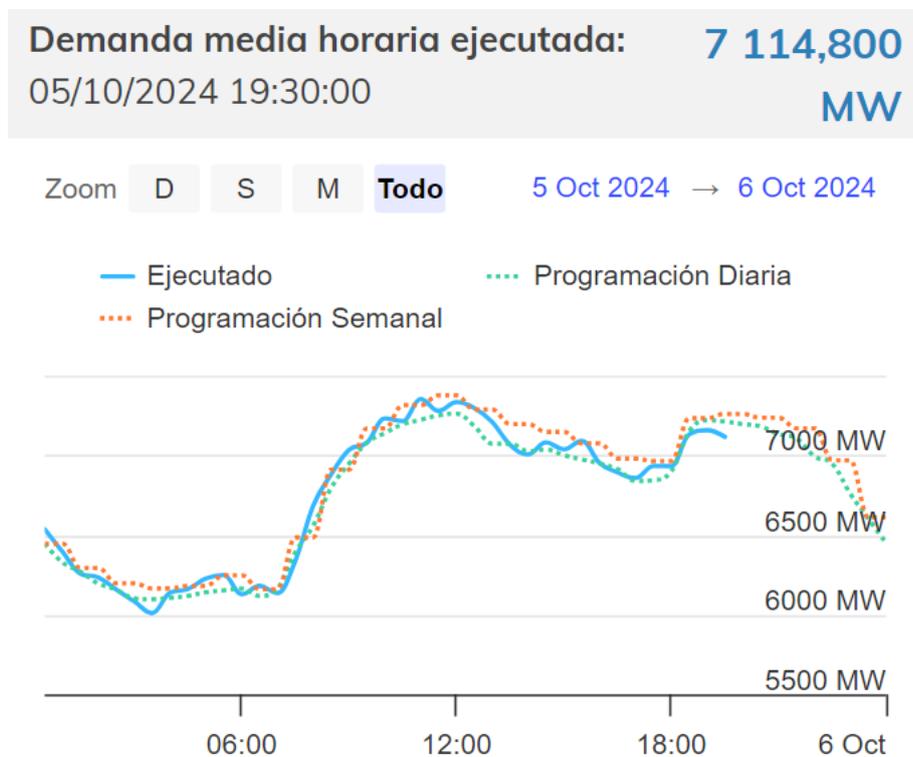
La demanda de energía eléctrica varía a lo largo del día, la semana y el año,

influenciada por factores como los patrones de consumo diario, las estaciones del año, la actividad económica y otros elementos (Osinergmin, 2017).

En Perú la demanda de energía eléctrica se caracteriza por registrar un comportamiento variable durante el día (ver Figura 2.3) donde se puede ver las tres demandas: demanda ejecutada, demanda de programación diaria y demanda de programación semanal.

Figura 2.3

Demanda de energía eléctrica en el Perú.



Fuente: (COES, 2024).

2.2.4. Pandemia COVID-19 en Perú

De acuerdo con la RAE, pandemia es una enfermedad epidémica que se extiende a muchos países o que ataca a casi todos los individuos de una localidad o región.

Por su parte la OMS, COVID-19 es una enfermedad infecciosa causada por el virus SARS-CoV-2 (OMS, 2020).

El brote de COVID-19, reportado por primera vez en enero de 2020 en Wuhan, China, rápidamente se convirtió en una pandemia global. Para el 29 de noviembre de 2020, se notificaron más de 61 millones de casos y 1.4 millones de muertes en todo el mundo. En las Américas, desde enero, se habían reportado más de 26 millones de casos. En Perú, el 6 de marzo de 2020 el presidente Martín Vizcarra confirmó el primer caso de infección por coronavirus COVID-19 en un joven procedente de Europa (EC, 2020).

A partir del 15 de marzo de 2020, el presidente de Perú declaró el estado de emergencia nacional y el aislamiento social obligatorio por un periodo inicial de 15 días, que luego se extendió durante varios meses (EP, 2020).

En octubre de 2022, el gobierno oficializó el fin del estado de emergencia por la COVID-19, basándose en la evaluación del contexto actual, que consideró la evolución de la pandemia, el avance del proceso de vacunación, la disminución de la tasa de positividad, la reducción de pacientes en las unidades de cuidados intensivos y la disminución de fallecimientos (MINSA, 2022).

Figura 2.4
Coronavirus en el Perú.



Fuente: (MINSA, 2022).

2.2.5. Red neuronal LSTM

Las LSTM (*Long Short-Term Memory*) son un tipo de red neuronal recurrente (RNN) diseñada para resolver el problema conocido como desvanecimiento del gradiente (*vanishing gradient problem*), el cual enfrentan las RNN tradicionales al intentar aprender dependencias a largo plazo. Las LSTM son particularmente eficaces para el procesamiento y la predicción de secuencias de datos debido a su capacidad para retener información relevante durante largos intervalos de tiempo (Hochreiter and Schmidhuber, 1997).

La estructura básica de una LSTM incluye una celda de memoria que puede mantener información a lo largo del tiempo, y tres tipos de puertas (Zhang et al., 2007):

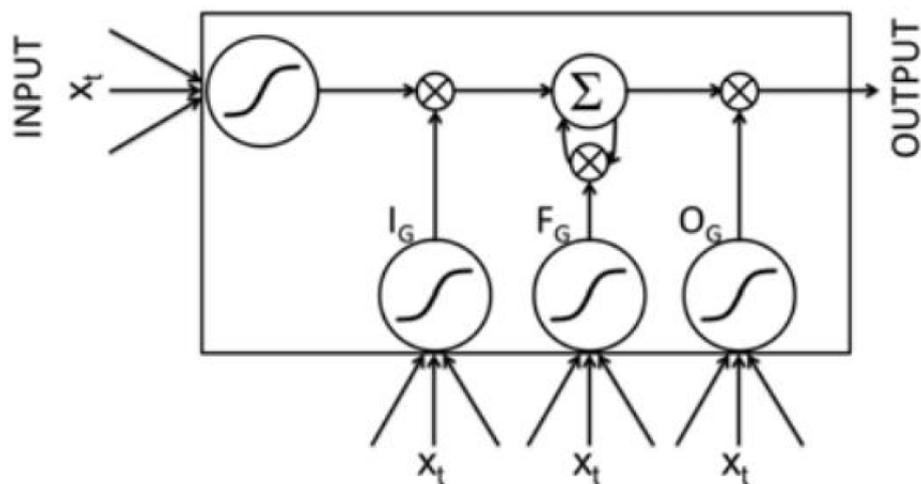
- Puerta de entrada: controla qué nueva información debe almacenarse en la celda de memoria.
- Puerta de olvido: decide qué parte de la información almacenada debe olvidarse.
- Puerta de salida: regula la cantidad de información que se envía como salida de la celda.

Estas puertas permiten a las LSTM manejar dependencias a largo plazo de manera más eficaz que las RNN simples, haciéndolas útiles en tareas como la predicción de series temporales, traducción automática y reconocimiento de voz (GeeksforGeeks, 2025).

Las redes neuronales LSTM son particularmente útiles para el aprendizaje de secuencias que contienen patrones a largo plazo de longitud desconocida, debido a su capacidad para mantener la memoria a largo plazo. El apilamiento de capas ocultas recurrentes en dichas redes también permite el aprendizaje de características temporales de nivel superior, para un aprendizaje más rápido con representaciones más dispersas (Malhotra et al., 2015) .

Las redes neuronales LSTM superan el problema del desvanecimiento del gradiente experimentado por las redes neuronales recurrentes (RNN) al emplear puertas multiplicativas que imponen un flujo de error constante a través de los estados internos de unidades especiales llamadas celdas de memoria. Las puertas de entrada (IG), salida (OG) y olvido (FG) evitan que los contenidos de la memoria se vean perturbados por entradas y salidas irrelevantes (ver Figura 2.5), lo que permite el almacenamiento de memoria a largo plazo. Debido a esta capacidad de aprender correlaciones a largo plazo en una secuencia, las redes LSTM obvian la necesidad de una ventana de tiempo preespecificada y son capaces de modelar con precisión secuencias multivariadas complejas (Malhotra et al., 2015).

Figura 2.5
Célula de LSTM.



Fuente: (Malhotra et al., 2015).

2.3. Antecedentes Empíricos de la Investigación

2.3.1. Dinh et al. (2024), *Electrical Energy Demand Forecasting using Time Series in LSTM and CNN-LSTM Models in Deep Learning Applications*, Electrical Energy Demand Forecasting using Time Series in LSTM and CNN-LSTM, USA.

El estudio tuvo como objetivo evaluar el desempeño de modelos de aprendizaje profundo para la predicción de la demanda eléctrica utilizando series temporales univariadas. Se implementaron dos enfoques: una red Bidireccional LSTM (BiLSTM) y un modelo híbrido CNN-LSTM, ambos aplicados a un conjunto de datos públicos de operadores de servicios de transmisión (TSO). Previamente al entrenamiento, los datos (2015–2018, con registros horarios) fueron limpiados, normalizados y transformados en secuencias de entrada y salida. Los modelos se entrenaron con hiperparámetros optimizados (capas LSTM con 64 unidades, regularización L2, dropout y función de pérdida MSE) y se evaluaron utilizando métricas como MAE, MAPE, RMSE y R^2 . El modelo CNN-LSTM incluyó capas convolucionales y de pooling para extraer características espaciales antes de las capas LSTM. Los resultados mostraron que el modelo híbrido CNN-LSTM superó al BiLSTM en todas las métricas: MAE de 499.08 frente a 780.56, MAPE de 1.80 % frente a 2.52 %, RMSE de 671.37 frente a 1042.20 y R^2 de 0.97 frente a 0.94. Se concluye que la combinación de CNN y LSTM mejora la precisión en la predicción de series temporales univariadas, siendo el modelo CNN-LSTM una alternativa robusta para la gestión eficiente de la demanda energética. Este artículo respalda la eficacia de modelos LSTM y CNN-LSTM en la predicción de demanda energética, validando el enfoque metodológico de esta investigación.

2.3.2. Rahman et al. (2024), *Time Series-Based Analysis of Energy Consumption: Forecasting and Anomaly Detection Using LSTM and Isolation Forest*, International Journal of Engineering Research and Technology, USA.

Tuvo como objetivo desarrollar un sistema integrado que permita predecir el consumo energético residencial y detectar anomalías de forma automática, contribuyendo a una gestión más eficiente y segura del uso de la energía. Para ello, se utilizó una metodología basada en series temporales que incluyó etapas de preprocesamiento de datos (limpieza, normalización y resampling diario), análisis exploratorio visual para identificar tendencias y estacionalidades, y la aplicación de dos modelos principales: una red neuronal LSTM para la predicción de consumo y el algoritmo Isolation Forest para la detección de anomalías. La evaluación del modelo LSTM se realizó mediante la métrica RMSE, mientras que la detección de anomalías se validó con precisión, recall y F1-score. Los resultados mostraron que el modelo LSTM alcanzó un RMSE de 0.0799, lo que indica una alta precisión en las predicciones. Por su parte, el modelo Isolation Forest identificó correctamente el 100% de las anomalías (258 casos), con valores perfectos de precisión, recall y F1-score (1.0), sin falsos positivos ni negativos. Estos hallazgos evidencian la eficacia del enfoque combinado para aplicaciones reales en el monitoreo y control del consumo energético. Este artículo respaldó el uso de redes LSTM para predecir con alta precisión el consumo energético y mostró que su combinación con Isolation Forest mejoró la detección automática de anomalías. Validó el enfoque no supervisado adoptado en el presente estudio. Además, ofreció una base sólida para futuras mejoras híbridas en el monitoreo energético.

2.3.3. Hanif et al. (2021), *Power Consumption Predictive Analytics and Automatic Anomaly Detection Using CNN-LSTM*, 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE).

El estudio tuvo como objetivo desarrollar un modelo predictivo de consumo energético y detección automática de anomalías utilizando una arquitectura híbrida CNN-LSTM, a fin de mejorar la precisión en la predicción y promover el uso eficiente de la energía. Se recolectaron datos eléctricos mediante un módulo EMS doméstico que registró información cada minuto, la cual fue posteriormente remuestreada a intervalos horarios y analizada con técnicas estadísticas y visuales para identificar outliers y patrones. El modelo propuesto combinó una red convolucional (CNN) para extraer características relevantes con una red LSTM para capturar dependencias temporales, y fue entrenado con datos supervisados generados por una ventana deslizante de 24 horas. Se utilizó el error cuadrático medio (MSE) como métrica principal y el optimizador Adam en un entorno de entrenamiento de 50 épocas. La detección de anomalías se basó en la comparación entre valores reales y predichos mediante el cálculo del MAE, y se estableció un umbral automático para clasificar valores anómalos. Los resultados mostraron que el modelo CNN-LSTM superó al LSTM estándar, reduciendo el MSE en un 29% (2097 vs. 2713), y demostró eficacia al detectar apagones y consumos inusuales, validando su aplicabilidad en la monitorización inteligente del consumo eléctrico. Este estudio demostró que la arquitectura híbrida CNN-LSTM mejoró la precisión en la predicción energética y la detección de anomalías frente al modelo LSTM estándar. Validó la utilidad de comparar valores reales y predichos para identificar eventos anómalos, tal como se aplicó en esta tesis. Además, ofreció un enfoque complementario que podría adoptarse como mejora futura del modelo.

2.3.4. Blázquez G. et al. (2021), *A review on outlier/anomaly detection in time series data*, ACM computing surveys (CSUR), USA.

La revisión de técnicas de detección de outliers presentada en este artículo ofrece una clasificación organizada que facilita la comparación y selección de métodos según el tipo de datos y la naturaleza del outlier. Sin embargo, la ambigüedad en la terminología donde los términos outlier, discord y anomaly se utilizan de manera variable, resalta la necesidad de un lenguaje estandarizado en la literatura, lo cual es esencial para evitar malentendidos. Además, al considerar los outliers como eventos de interés y no simplemente como datos a eliminar, se abre la posibilidad de extraer información valiosa de ellos. A pesar de que la mayoría de los enfoques actuales se centran en series temporales univariantes, es crucial desarrollar métodos que aborden la complejidad de las series multivariantes. Por último, se debe tener cuidado con los conjuntos de referencia utilizados para determinar el valor esperado, ya que la inclusión de outliers en estos datos puede distorsionar los resultados. En conjunto, esta revisión sienta las bases para futuros desarrollos en la detección de anomalías en series temporales, promoviendo un enfoque más crítico y proactivo en el análisis de datos.

2.3.5. Ergen and Kozat (2019), *Unsupervised and Semi-supervised Anomaly Detection with LSTM Neural Networks*, IEEE transactions on neural networks and learning systems, USA.

El artículo presenta avances significativos en la detección de anomalías mediante una estructura LSTM genérica diseñada para manejar secuencias de longitud variable, lo que proporciona una mayor flexibilidad y capacidad de adaptación para abordar la

diversidad de series temporales. La introducción de un método de optimización conjunta que ajusta los parámetros de la arquitectura LSTM y de funciones de puntuación como OC-SVM y SVDD mejora notablemente la efectividad del modelo, reflejando la importancia de la personalización en la detección de anomalías. Los resultados obtenidos muestran que los algoritmos propuestos superan a los métodos convencionales, evidenciando su superioridad tanto en conjuntos de datos simulados como reales, lo que refuerza la validez de combinar LSTM con técnicas de detección de anomalías. Además, la implementación de métodos de entrenamiento basados en programación cuadrática y gradiente amplía la aplicabilidad de estos algoritmos a contextos semi-supervisados y supervisados, ofreciendo un enfoque más robusto para la formación de modelos. Estos desarrollos son particularmente relevantes para el análisis de la demanda de energía eléctrica, ya que la capacidad de procesar secuencias de longitud variable se alinea con las características complejas de los datos que se utilizarán en la investigación, lo que sugiere que estas innovaciones podrían mejorar significativamente la precisión en la identificación de patrones irregulares en dicho contexto.

2.3.6. Niu et al. (2020), *LSTM-Based VAE-GAN for Time-Series Anomaly Detection*, Sensors 2020, China.

El artículo presenta un enfoque innovador para la detección de anomalías mediante la combinación de VAE-GAN y LSTM, lo que resalta su relevancia y aplicabilidad en el monitoreo de estados de equipos a través de series temporales de datos en entornos industriales. Este método se estructura en dos etapas: la primera, dedicada al entrenamiento del modelo para aprender la distribución de datos normales, y la segunda, centrada en la detección de anomalías a través de puntuaciones específicas. Los experimentos realizados indican que el método VAE-GAN no solo supera a varios enfoques clásicos en términos de valor F1 en la detección de anomalías, sino que también optimiza el proceso de detección para mejorar la velocidad y precisión, lo que es esencial en aplicaciones industriales donde el tiempo de respuesta es crítico.

Esta combinación de técnicas ofrece un marco teórico y práctico que no solo fortalece el modelado y validación de redes neuronales LSTM, sino que también facilita su implementación en el análisis de series temporales, haciendo hincapié en su potencial para mejorar la eficiencia en la detección de patrones irregulares. Este artículo contribuyó significativamente al modelado, adaptación y validación de la red neuronal LSTM para su aplicación en series temporales. Proporciona un marco que facilitará la implementación de LSTM.

2.3.7. Bakhtawar M. (2019), *Anomaly detection in electricity demand time series data*, Kth Royal Institute of Technology School of Electrical Engineering and Computer Science, Suecia.

Esta tesis destaca la superioridad del modelo LSTM sobre el modelo de red neuronal feedforward (FFNN) en la detección de patrones en series temporales con regularidades diarias y semanales, lo que sugiere que el LSTM es más adecuado para capturar dependencias temporales en datos predecibles, como los de la demanda eléctrica. Sin embargo, la falta de diferencias significativas en el rendimiento entre ambos modelos en conjuntos de datos sin patrones evidentes indica que la eficacia de los modelos depende de la naturaleza de los datos, lo que plantea desafíos en situaciones con ruido o sin patrones claros. Además, los resultados subrayan la necesidad de cautela al interpretar las diferencias de rendimiento, ya que la falta de significancia estadística limita la afirmación de la superioridad de un modelo sobre otro, resaltando la importancia de realizar más investigaciones para identificar el modelo más efectivo en una variedad más amplia de conjuntos de datos. En este contexto, la tesis ayudó en la aplicación del modelo LSTM para la detección de anomalías y proporciona orientación sobre cómo adaptar y modelar estos algoritmos, lo cual será fundamental para analizar la demanda de energía durante y después de la pandemia, donde los datos pueden

variar considerablemente, especialmente en la configuración de hiperparámetros del modelo.

2.3.8. Malhotra et al. (2015), *Long Short Term Memory Networks for Anomaly Detection in Time Series*, ES-SANN 2015 23rd European Symposium on Artificial Neural Networks Computational Intelligence and Machine Learning, Belgica.

El artículo revela que las redes LSTM apiladas son altamente efectivas para la detección de patrones a largo plazo en series temporales, lo que las convierte en herramientas poderosas para la identificación de anomalías. El hecho de que el modelo se entrene en datos no anómalos para predecir futuras observaciones y utilice la distribución gaussiana multivariada de los errores de predicción para detectar comportamientos inusuales representa un enfoque innovador y estadísticamente sólido para el análisis de datos. Sin embargo, aunque los resultados sugieren que el LSTM es más efectivo en la presencia de patrones regulares en los datos, la falta de significancia estadística impide afirmar la superioridad de este modelo en comparación con otros, lo que implica que la investigación debe continuar para explorar y validar su eficacia en diferentes contextos. La validación del enfoque en una variedad de conjuntos de datos, como el ECG y la demanda de energía, refuerza su versatilidad y aplicabilidad en diversas áreas de detección de fallos y anomalías. Este aspecto es especialmente relevante para el análisis de la demanda de energía eléctrica en el contexto pre y post-pandemia, donde la identificación de patrones irregulares puede ser crucial para la gestión y optimización del consumo energético. En resumen, el artículo proporcionó un marco valioso que puede ser aprovechado para mejorar la detección de anomalías en series temporales, destacando la relevancia de adaptar y aplicar estas técnicas a situaciones específicas.

2.3.9. Pan et al. (2022), *High-Dimensional Energy Consumption Anomaly Detection: A Deep Learning - Based Method for Detecting Anomalies*, School of Mechanical and Automatic, Zhejiang Sci - Tech University, Hangzhou, China.

Esta investigación aborda el problema de identificar comportamientos anómalos en el consumo eléctrico a partir de datos de alta dimensión, dada la creciente demanda energética y la necesidad de reducir el desperdicio. El objetivo principal es desarrollar un método de detección de anomalías en tiempo real utilizando una combinación de redes neuronales convolucionales (CNN), redes LSTM bidireccionales (Bi-LSTM) y mecanismos de atención, aplicando el criterio estadístico de 3 desviaciones estándar para clasificar eventos como anómalos. Los resultados, validados con un conjunto de datos del UCI sobre consumo energético en viviendas, muestran que la combinación propuesta mejora significativamente la precisión (MAE y MAPE) respecto a modelos que usan componentes aislados. Este enfoque permite predecir el consumo y detectar desviaciones en tiempo real, lo cual es útil tanto para gestores de edificios como para usuarios. Este trabajo proporcionó una base metodológica sólida para aplicar modelos LSTM en la detección de anomalías en series temporales de demanda eléctrica, facilitando el análisis del comportamiento energético durante y después de la pandemia.

2.3.10. Cui et al. (2023), *A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting*, Department of Computer Science, Changzhi University, Changzhi, Shanxi, China.

Este artículo aborda el impacto de la pandemia de COVID-19 sobre la demanda eléctrica y propone un marco de predicción basado en aprendizaje profundo ajustado a este contexto. El problema identificado es que la pandemia alteró drásticamente los patrones normales de consumo eléctrico, afectando negativamente la precisión de los modelos tradicionales de predicción. El objetivo del estudio fue diseñar un modelo que primero elimine el efecto de COVID-19 mediante regresión lineal con variables específicas (como confinamiento, fines de semana y estacionalidades), y luego predecir la demanda utilizando un modelo LSTM multiescalar y desestacionalizado (LSTM-MSNet-DS) sobre los residuos de esa regresión. Aplicado a datos reales de Taixing (China), los resultados muestran que el marco propuesto supera a otros siete modelos (como SVR, BiLSTM y GRU) en métricas de error (MAE, RMSE, MRE), logrando mejoras superiores al 17%. Este estudio ayuda directamente en la investigación, ya que ofrece una estrategia efectiva para modelar y pronosticar la demanda eléctrica en periodos anómalos como la pandemia, y puede ser adaptado para comparar las anomalías entre los periodos durante y post-COVID-19 en los datos de esta tesis.

Capítulo 3

Metodología

3.1. **Ámbito de estudio: Localización política y geográfica**

El ámbito de estudio se encuentra en el sector de la energía eléctrica, y estuvo enfocado en la detección de anomalías en series temporales mediante redes neuronales LSTM. El tema central de la investigación fue el análisis del impacto de la pandemia de COVID-19 en la demanda de energía eléctrica.

La localización política y geográfica de la investigación fue Perú, donde se analizaron datos de la demanda de energía eléctrica recopilados durante los últimos seis años.

3.2. **Tipo y nivel de investigación**

Esta investigación es de tipo aplicada, ya que busca resolver un problema específico del entorno real: la identificación de comportamientos anómalos en la demanda eléctrica peruana en contextos disruptivos como la pandemia de COVID-19. Su fina-

lidad es generar conocimiento útil y con impacto en la gestión energética, lo cual caracteriza a la investigación aplicada según (Hernández Sampieri et al., 2014).

El nivel de investigación es predictivo, pues se utiliza un modelo de redes neuronales LSTM entrenado con datos históricos para generar valores esperados de demanda, que posteriormente se comparan con los datos reales para detectar anomalías. Este nivel permite anticipar comportamientos futuros y contrastarlos con observaciones reales, un enfoque clave en estudios basados en series temporales.

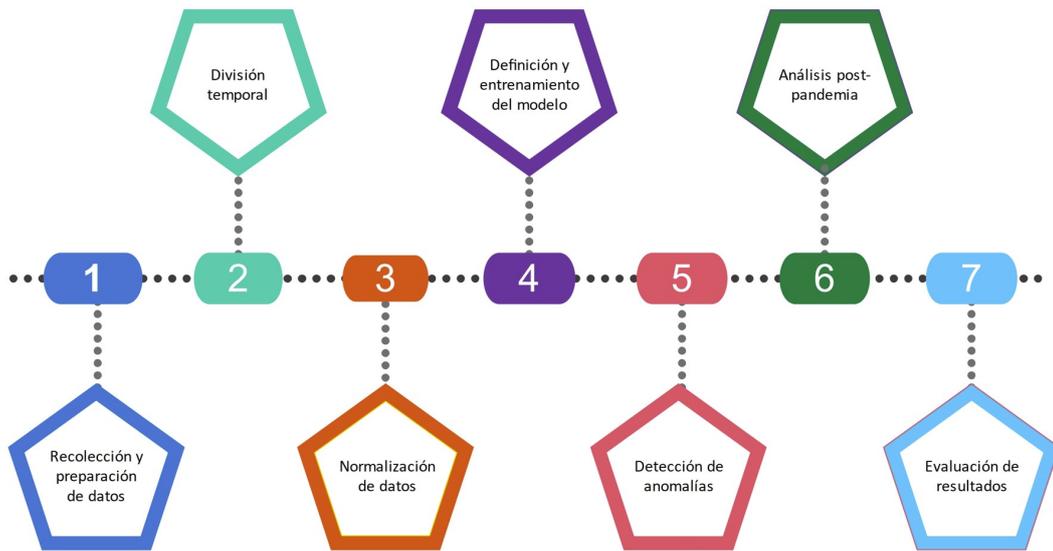
El enfoque metodológico es cuantitativo, ya que se utilizan herramientas matemáticas, estadísticas y computacionales para analizar datos numéricos (series de tiempo). Se aplica una técnica de aprendizaje automático supervisado mediante redes neuronales recurrentes del tipo LSTM (*Long Short-Term Memory*), que son ampliamente utilizadas en problemas de predicción y detección de patrones en series temporales.

3.3. Metodología de la investigación

La presente investigación sigue un enfoque cuantitativo, predictivo y aplicado, centrado en la modelación de series temporales para la detección de anomalías en la demanda de energía eléctrica en el Perú. Se empleó una metodología estructurada que abarcó la recopilación de datos históricos, el preprocesamiento, el entrenamiento de un modelo LSTM y la posterior detección de comportamientos atípicos mediante el análisis de errores.

A continuación, se presenta el esquema metodológico seguido en esta investigación (ver Figura 3.1).

Figura 3.1
Pasos metodología de investigación



Fuente:Elaboración propia.

1. **Recolección y preparación de datos:** Recopilar los datos históricos de demanda eléctrica de los últimos seis años, asegurándose de que estén limpios y bien organizados.
2. **División temporal:** Separar los datos en tres etapas: pre-pandemia, durante la pandemia y post-pandemia.
3. **Normalización de datos:** Normalizar los datos para facilitar el entrenamiento del modelo LSTM.
4. **Definición y entrenamiento del modelo LSTM:** Definir la capa de entrada, las capas intermedias para captar las dependencias temporales y la capa de salida. Se utilizarán los datos de la etapa pre-pandemia para entrenar el modelo y establecer patrones normales.
5. **Detección de anomalías:** Usar el modelo entrenado para predecir la demanda durante y después de la pandemia, calcular el error entre las predicciones y los datos reales, y así identificar posibles anomalías.

6. **Análisis post-pandemia:** Repetir el proceso de detección para analizar si las anomalías persisten después de la pandemia.
7. **Evaluación de resultados:** Interpretar las anomalías encontradas en relación con los cambios en la demanda, considerando el impacto de la pandemia.

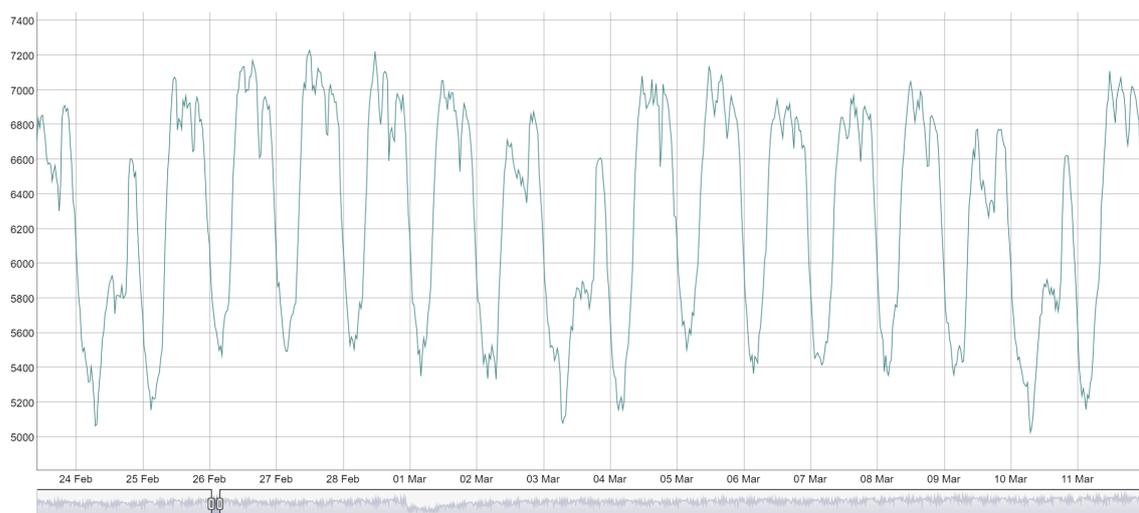
3.4. Unidad de análisis

La unidad de análisis fue la serie temporal de la demanda de energía eléctrica, un dato cuantitativo continuo, medido en MW (megavatios).

En la Figura 3.2 se muestra la serie temporal de la demanda de energía eléctrica durante un periodo de dos semanas. A simple vista, se aprecia un comportamiento cíclico con picos característicos en las horas de media mañana y noche, así como una mayor variabilidad en el consumo nocturno. También se evidencia una diferencia notable entre los días hábiles y los fines de semana.

Figura 3.2

Demanda de energía eléctrica en Perú: lapso de 2 semanas.



Fuente: Elaboración propia con RSTUDIO.

3.5. Población de estudio

En este estudio, la población de interés está constituida por todos los datos de demanda de energía eléctrica registrados cada 30 minutos desde el 14 de marzo de 2018 hasta el 13 de marzo de 2024, con un total de 105,216 observaciones. Este rango temporal abarca tres etapas específicas: el periodo anterior a la pandemia de COVID-19, los dos años durante la pandemia y el periodo post pandemia. Los datos fueron extraídos de la página oficial del Comité de Operación Económica del Sistema Interconectado (COES), disponible en <https://www.coes.org.pe/Portal/>.

3.6. Técnicas de recolección de información

Es una técnica **documental**, por que los registros existen y pertenecen a la información de una base de datos. Según la Técnica para el Intercambio de Información en Tiempo Real para la Operación del Sistema Eléctrico Interconectado Nacional Resolución Directoral N° 243-2012-E-DGE, donde en su artículo quinto indica que: por consideraciones de seguridad nacional, el COES deberá considerar que la eventual publicación de información en tiempo real o cuasi real. Por ende los datos obtenidos de tal página son confiables.

3.7. Métricas para evaluar el modelo de regresión

3.7.1. Raíz del error cuadrático medio

RMSE (*Root Mean Square Error*) mide la raíz cuadrada del error cuadrático medio entre los valores reales y las predicciones del modelo.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Donde y_i son los valores reales y \hat{y}_i son los valores predichos.

Esta métrica penaliza los errores grandes porque eleva al cuadrado las diferencias y su resultado siempre está en las mismas unidades de los datos originales. Es útil cuando se quiere detectar si hay errores grandes en las predicciones (Montgomery et al., 2021).

3.7.2. Error absoluto medio

El error absoluto medio (MAE, *Mean Absolute Error*) mide el promedio de los errores absolutos entre los valores reales y predichos.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

Donde y_i son los valores reales y \hat{y}_i son los valores predichos.

Esta métrica es robusta a valores atípicos porque no eleva al cuadrado los errores, y el resultado está en las mismas unidades de los datos originales (Montgomery et al., 2021).

3.7.3. Coeficiente de determinación

El coeficiente de determinación (R^2) mide qué porción de la variabilidad en los datos reales es explicada por el modelo.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Donde \bar{y} es la media de los valores reales.

Este puede tomar valores entre $[0,1]$, donde:

- $R^2 = 1$, el modelo predice perfectamente los valores.
- $R^2 = 0$, el modelo no explica la variabilidad de los datos.
- $0 < R^2 < 1$, el modelo se ajusta medianamente o parcialmente a los datos.

Esta métrica sirve para evaluar la bondad de ajuste del modelo e indica si el modelo captura la variación en los datos (Montgomery et al., 2021).

En conclusión se utilizó estas 3 métricas porque:

- RMSE: Ideal para modelos de predicción en series temporales donde es crucial minimizar grandes errores.
- MAE: Es más interpretativo y adecuado si no quieres penalizar excesivamente los valores atípicos.
- R^2 : Da una visión general de qué tan bien el modelo captura la variabilidad de los datos.

3.8. Herramientas informáticas utilizadas

Las herramientas informáticas que se utilizaron en esta investigación fueron:

- Entorno de desarrollo: El modelo fue desarrollado y ejecutado en Google Colaboratory, un entorno en la nube que permite trabajar con notebooks de Python y utilizar recursos computacionales como GPU de forma gratuita.
- Python: Se utilizó Python para implementar redes neuronales LSTM con el objetivo de detectar y analizar anomalías en la demanda de energía eléctrica en diferentes periodos de la pandemia de COVID-19.
 - Versión: 3.11.12
 - Librerías: pandas 2.2.2, tensorflow 2.18.0, matplotlib 3.10.0.
- R: RStudio se utilizó para la generación de gráficos y el análisis estadístico de los datos, permitió una exploración detallada, visualización de de patrones, y características en la serie temporal antes, durante y después de la pandemia.
 - Versión: 2025.05.0 Build 496
 - R: 4.4.2
 - Librerías: dygraphs 1.1.1.6, ggplot2 3.5.1.

Capítulo 4

Resultados

A continuación, se describen en detalle los pasos correspondientes a cada etapa de la metodología de investigación.

4.1. Recolección y preparación de datos

4.1.1. Acceso y recolección de datos

Los datos se obtuvieron del Comité de Operación Económica del Sistema Interconectado Nacional (COES, <https://www.coes.org.pe/portal/>). El archivo descargado se muestra en la Figura 4.1.

Figura 4.1

Archivo de demanda de energía eléctrica.

	A	B	C	D
1				
2				
3				
4	FECHA	EJECUTADO	PROG. DIARIA	PROG. SEMANAL
5	14/03/2018 00:30	5520.10984	5493.71448	5476.48661
6	14/03/2018 01:00	5431.98302	5382.76515	5476.48661
7	14/03/2018 01:30	5326.25206	5319.82871	5327.61256
8	14/03/2018 02:00	5254.40885	5262.05276	5327.61256
9	14/03/2018 02:30	5162.80993	5211.19157	5265.09237
10	14/03/2018 03:00	5139.37646	5174.12462	5265.09237
11	14/03/2018 03:30	5141.75917	5174.2	5227.47547
12	14/03/2018 04:00	5215.52212	5170.52597	5227.47547
13	14/03/2018 04:30	5232.62383	5199.29125	5286.41983
14	14/03/2018 05:00	5293.11045	5248.34011	5286.41983

Fuente: (COES, 2024).

4.1.2. Descripción del archivo de datos

El archivo descargado está en formato .xlsx y contiene cuatro columnas, las cuales se describen a continuación:

- ****Fecha****: Corresponde a la fecha y hora de la demanda de energía eléctrica, con el formato DD/MM/AAAA HH:MM.
- ****Ejecutado****: Valor numérico correspondiente a la demanda de energía eléctrica consumida, en megavatios (MW).
- ****Prog. diaria****: Valor numérico de la demanda de energía eléctrica proyectada diariamente, en megavatios (MW).
- ****Prog. semanal****: Valor numérico de la demanda de energía eléctrica proyectada semanalmente, en megavatios (MW).

4.1.3. Descripción general de los datos

Se trabajó con las dos primeras columnas: **fecha** y **ejecutado**. Las fechas seleccionadas abarcan desde el 14 de marzo de 2018 al 13 de marzo de 2024, con un registro cada 30 minutos, sumando un total de 105,216 observaciones. Para facilitar el tratamiento de los datos, el archivo fue convertido a formato .csv.

4.1.4. Exploración inicial

En esta etapa inicial, se realizó una prueba para garantizar la periodicidad de los datos en la columna **fecha**, mientras que en las demás columnas se verificó la existencia de datos nulos o faltantes. La periodicidad y la completitud de todas las columnas se validaron exitosamente.

Inicialmente, se verificó que los tipos de datos eran *object* y *float64* para las columnas **fecha** y **ejecutado**, respectivamente. Posteriormente, la columna **fecha** se convirtió al formato *datetime64*.

Tabla 4.1

Estadísticas descriptivas de la variable EJECUTADO.

Estadística	Ejecutado
Cantidad	105,216.000000
Media	6,224.810796
Desviación estándar	730.501800
Mínimo	3,249.729290
25 %	5,728.518845
Mediana	6,291.400745
75 %	6,752.305710
Máximo	8,304.612840

En la Tabla 4.1 muestra un resumen de estadísticas descriptivas de la variable en estudio que es la serie temporal de demanda de energía eléctrica en Perú, con 105,216 observaciones, representando una serie temporal de tamaño considerable. El promedio es de 6,224.81 MW y la desviación estándar es de 730.50 MW, lo cual indica

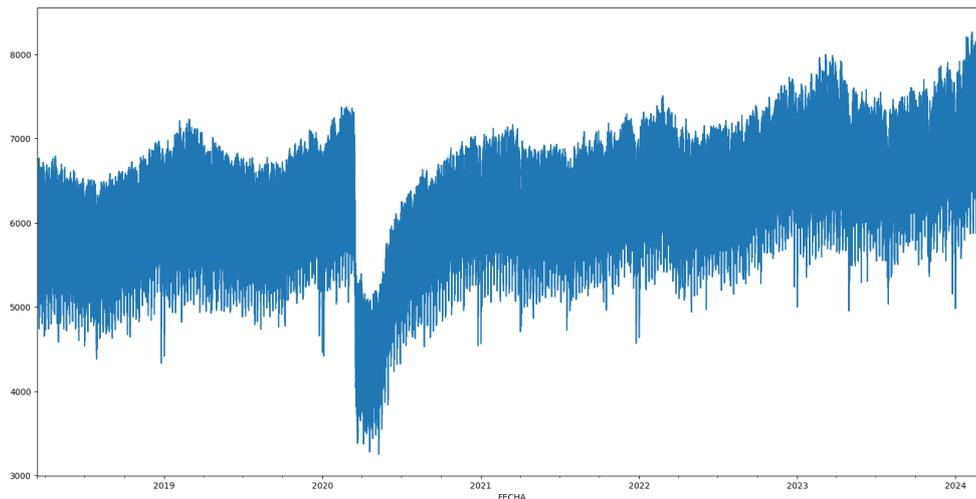
una dispersión moderada respecto al promedio. El valor mínimo es de 3,249.72 MW y el máximo de 8,304.61 MW, mientras que la mediana es de 6,291.40 MW.

4.1.5. Análisis de gráficos iniciales

A continuación, en la Figura 4.2, se presenta la serie temporal completa del consumo de energía eléctrica. Visualmente, se observa una disminución considerable en la demanda durante el periodo de inicio de la emergencia nacional y el aislamiento social obligatorio debido a la pandemia de COVID-19, con una recuperación gradual a partir de inicios del año 2021.

Figura 4.2

Demanda de energía eléctrica en Perú 2018 - 2024.



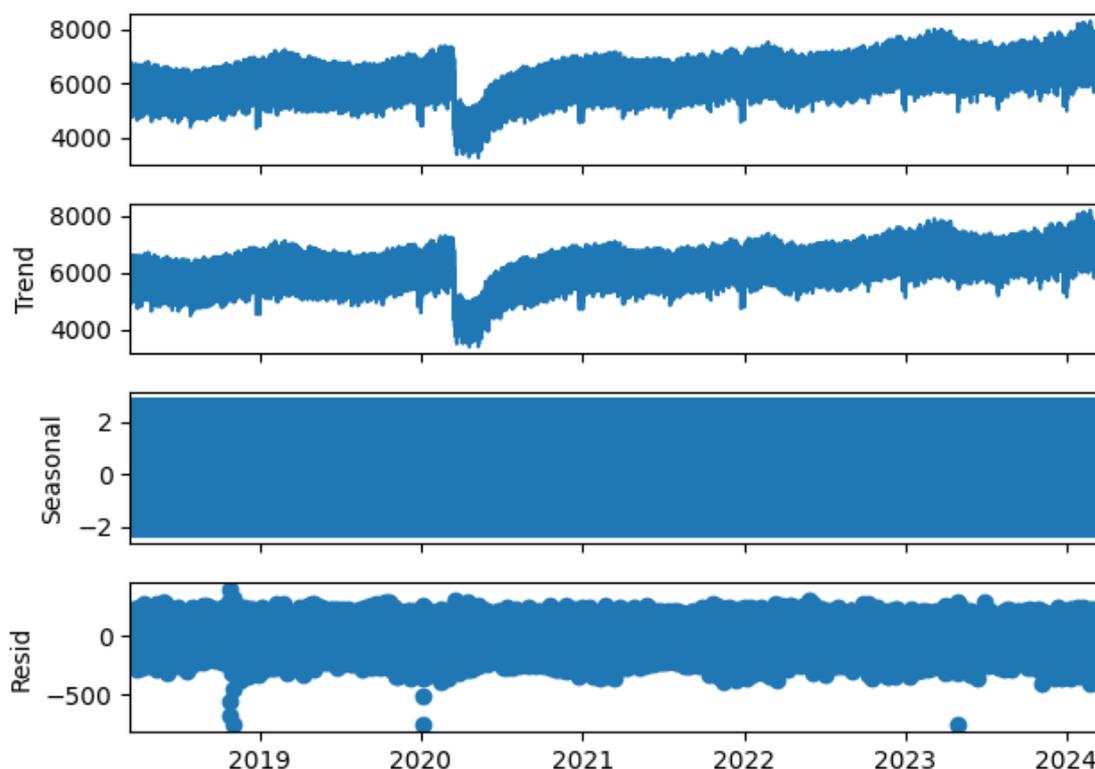
Fuente: Elaboración propia.

Para entender de manera inicial la serie temporal, se descompuso en sus componentes principales (ver Figura 4.3). En la serie original se observa una disminución notable alrededor de 2020, coincidiendo claramente con el inicio de la pandemia en Perú. La tendencia (*trend*) muestra el componente de largo plazo eliminando las fluctuaciones de corto plazo; aquí, la tendencia presenta una caída marcada en 2020, seguida de un incremento constante que refleja una recuperación paulatina hacia los niveles originales. La estacionalidad (*seasonal*) muestra patrones repetitivos regulares,

que incluyen variaciones diarias, semanales y anuales, con una amplitud relativamente pequeña comparada con los demás componentes. Por otro lado, los residuos (*resid*) representan fluctuaciones no explicadas por la tendencia o estacionalidad, siendo más pronunciados alrededor de 2020, lo que sugiere eventos anómalos o cambios bruscos no explicados por los otros componentes.

Figura 4.3

Descomposición de la serie temporal.



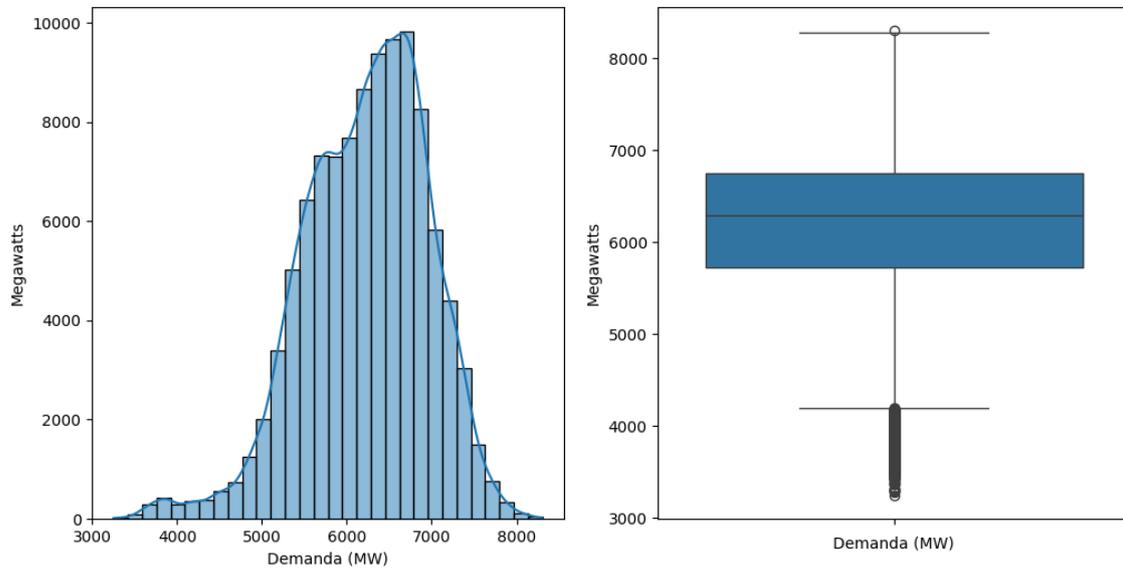
Fuente: Elaboración propia.

En la Figura 4.4 se presentan dos representaciones de la demanda de energía eléctrica en Perú: un histograma (izquierda) y un diagrama de caja (derecha). En el histograma, la distribución es unimodal con un sesgo leve hacia la izquierda, concentrando la mayoría de los valores entre 6,000 y 7,000 MW, lo que indica que estas son las demandas más frecuentes. El diagrama de caja refuerza esta observación, con un rango intercuartílico entre aproximadamente 6,000 y 7,000 MW y una mediana cercana a 6,500 MW. Sin embargo, se observan valores atípicos significativos por debajo de 4,000 MW, los cuales podrían representar eventos inusuales o condiciones extremas en

la demanda eléctrica.

Figura 4.4

Histograma y diagrama de caja de la demanda de energía eléctrica.



Fuente: Elaboración propia.

Esta etapa inicial de recolección y preparación de datos permitió consolidar una base de 105,216 registros con periodicidad de 30 minutos, garantizando la calidad y completitud de la información. Los análisis preliminares revelaron una demanda promedio de 6,224.81 MW, con una desviación estándar de 730.50 MW, indicando una variabilidad moderada en los datos. Además, la descomposición de la serie temporal destacó cambios significativos en la tendencia durante el inicio de la pandemia, así como patrones estacionales regulares y fluctuaciones residuales más pronunciadas en 2020. Estos hallazgos sugieren que las anomalías podrían estar asociadas a eventos disruptivos específicos, como la pandemia, y proporcionan una referencia inicial para entrenar el modelo predictivo en los siguientes pasos. Se identificaron valores atípicos en los rangos más bajos, lo que refuerza la necesidad de un análisis cuidadoso de las anomalías durante las fases posteriores del estudio.

4.2. División temporal

4.2.1. Definición de las etapas

La serie temporal de demanda de energía eléctrica se dividió en tres etapas: dos años antes de la pandemia, dos años durante la pandemia y dos años posteriores a la pandemia. Los rangos de fechas correspondientes se detallan a continuación.

- Pre-pandemia: Del 14 de marzo de 2018 al 13 de marzo de 2020.
- Durante la pandemia: Del 14 de marzo de 2020 al 13 de marzo de 2022.
- Post-pandemia: Del 14 de marzo de 2022 al 13 de marzo de 2024.

La Figura 4.5 muestra la evolución de la demanda de energía eléctrica (en megavatios, MW) a lo largo del periodo comprendido entre 2018 y 2024, segmentada en tres etapas clave: pre-pandemia, durante la pandemia y post-pandemia. Cada fase se representa con un color distinto: azul para el periodo pre-pandemia (2018–inicio de 2020), naranja para el periodo durante la pandemia (marzo de 2020–marzo de 2022), y verde para el periodo post-pandemia (desde marzo de 2022 en adelante).

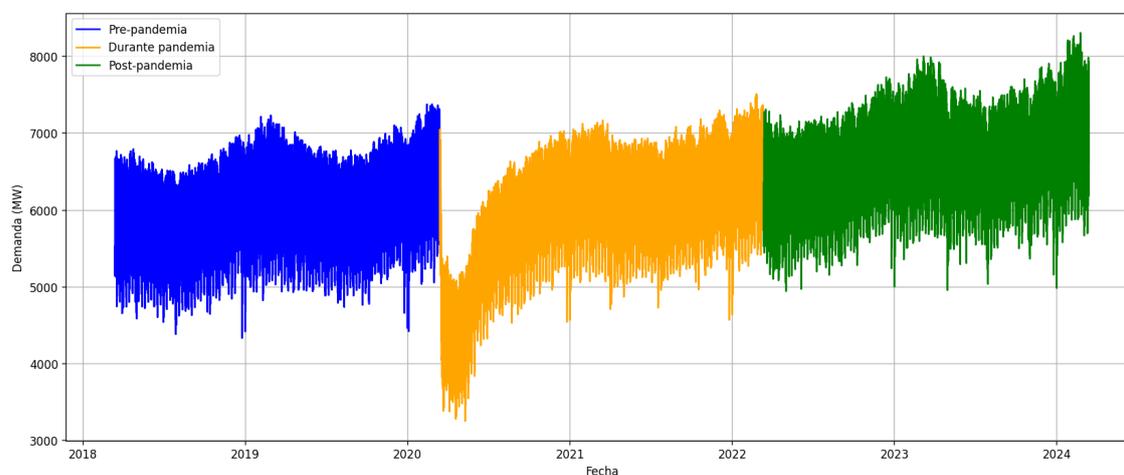
En la etapa pre-pandemia, se observa una demanda relativamente estable, con variaciones cíclicas diarias y semanales. A partir del inicio de la pandemia, representado por el cambio al color naranja, se aprecia una caída abrupta en los niveles de demanda, asociada al inicio del estado de emergencia y el aislamiento social obligatorio. Durante este periodo, la recuperación fue gradual pero inestable, con menor volumen de demanda y mayor dispersión en los datos, lo que sugiere alteraciones significativas en los patrones de consumo eléctrico.

Finalmente, en el periodo post-pandemia, coloreado en verde, se observa una recuperación más consistente y sostenida en la demanda de energía eléctrica. Aunque los valores aún presentan cierta variabilidad, se evidencia una tendencia al alza

en comparación con el periodo pandémico, lo cual podría indicar una normalización progresiva de las actividades económicas y sociales.

Figura 4.5

Etapas de la serie temporal del estudio



Fuente: Elaboración propia.

La decisión de tomar las ventanas de 2 años durante el periodo de pandemia, es porque, aunque las medidas de confinamiento comenzaron en marzo de 2020 y se mantuvieron formalmente hasta mediados de 2022, la población peruana empezó a relajar el cumplimiento de estas restricciones antes de su finalización oficial. En el diario internacional apnews.com, pese a que Perú impuso uno de los confinamientos más estrictos de América Latina, (*Many Peruvians ignore new virus lockdown orders*) debido al desgaste social y económico. Esto significa que, hacia los últimos meses del periodo, el comportamiento ciudadano ya no seguía el patrón impuesto por el gobierno, lo cual hizo sensible y representativa esta ventana temporal para capturar los efectos reales de la pandemia en la demanda energética, evitando datos contaminados por incumplimientos masivos.

4.2.2. Análisis estadístico por etapa

En la Tabla 4.2 y la Figura 4.6 se observan indicios de anomalías causadas por la pandemia. La dispersión en las etapas pre y post-pandemia es similar, pero

ambas difieren significativamente de la etapa durante la pandemia. Además, se pueden observar los valores máximos y mínimos de cada etapa, destacando que la etapa de la pandemia presenta una mayor amplitud en comparación con las otras dos. Asimismo, es evidente la presencia de datos atípicos durante la pandemia, mientras que las etapas pre y post-pandemia muestran distribuciones más homogéneas.

Tabla 4.2

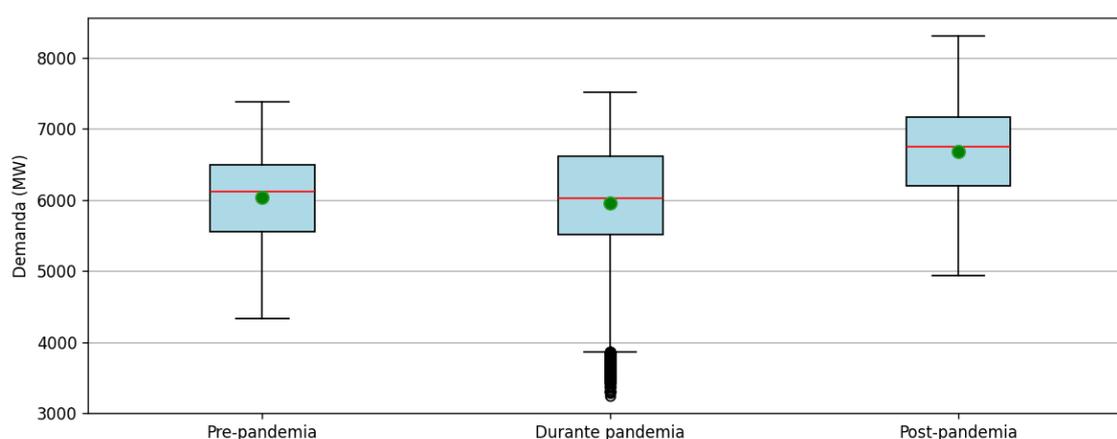
Resumen estadístico de la variable EJECUTADO en los tres periodos analizados.

Estadística	Pre-pandemia	Durante pandemia	Post-pandemia
Cantidad	35087.000000	34993.000000	35136.000000
media	6038.508455	5953.420752	6681.138837
std	568.571794	782.396436	590.999052
min	4330.632160	3249.729290	4940.226000
25 %	5552.578340	5510.010000	6194.428510
50 % (mediana)	6117.598330	6022.672600	6747.278930
75 %	6495.050615	6609.698290	7155.592722
max	7374.646060	7507.851730	8304.612840

La cantidad de datos por etapa es de 35,087, 34,993 y 35,136, respectivamente. Estas cifras no son iguales debido a que algunos años son bisiestos, lo que genera variaciones en la cantidad de registros.

Figura 4.6

Boxplot por etapa



Fuente: Elaboración propia.

La división de la serie temporal en tres etapas (pre-pandemia, durante la pandemia y post-pandemia) reveló diferencias significativas en el comportamiento de la

demanda de energía eléctrica. Durante la pandemia, se observó una mayor amplitud y presencia de valores atípicos en comparación con las etapas pre y post-pandemia, las cuales mostraron distribuciones más homogéneas y similares entre sí. Estadísticamente, la etapa de la pandemia tuvo la menor media (5,953.42 MW) y un mayor rango de dispersión, reflejando el impacto del aislamiento social y la emergencia sanitaria. Estos resultados confirman patrones anómalos específicos de la pandemia y una recuperación progresiva hacia la etapa post-pandemia.

4.2.3. Análisis estadístico inferencial de la demanda de energía eléctrica

Con el objetivo de respaldar posibles hallazgos obtenidos mediante el modelo LSTM, se realizó un análisis estadístico inferencial utilizando pruebas de t de Student, ANOVA y Tukey HSD para determinar si las diferencias en la demanda eléctrica entre los tres periodos analizados (pre-pandemia, durante la pandemia y post-pandemia) eran estadísticamente significativas.

A continuación se describe brevemente cada una de las pruebas en que consiste:

- Prueba t de Student (t-test): Es una prueba estadística que compara las medias de dos grupos independientes para determinar si existe una diferencia significativa entre ellas. Se usa para evaluar si los cambios observados en la media de un grupo respecto a otro pueden atribuirse al azar o son estadísticamente significativos (Montgomery and Runger, 2014). Para este estudio se comparó en grupo de 2 cada una de las etapas de la serie temporal de demanda de energía eléctrica.
- Análisis de Varianza (ANOVA): Es una técnica estadística que permite comparar las medias de tres o más grupos simultáneamente para verificar si al menos una difiere de las demás. Sirve como una prueba general para detectar diferencias significativas entre grupos (Field, 2013). Para esta investigación se comparó las

3 etapas.

- Prueba de comparaciones múltiples de Tukey (Tukey HSD): Es una prueba post hoc que se aplica después de un ANOVA significativo para identificar cuáles pares específicos de grupos presentan diferencias estadísticamente significativas. Se usa para realizar comparaciones múltiples controlando el error familiar. Es útil cuando hay más de dos grupos (Tukey, 1977). Se aplicó esta prueba para determinar las diferencias significativas entre fases temporales.

Lectura y preparación de los datos

Se utilizó el archivo `data_combinada.csv`, el cual contiene las columnas `FECHA`, `EJECUTADO` (demanda de energía) y `Fase` (etapa temporal). El siguiente código en R fue utilizado para cargar y preparar los datos:

```
1 # Leer archivo CSV
2 data <- read.csv("data_combinada.csv")
3
4 # Convertir columna a factor
5 data$Fase <- as.factor(data$Fase)
```

Pruebas t-test entre fases

Se aplicaron comparaciones por pares entre las fases utilizando la prueba t de Student para muestras independientes.

- Pre-pandemia vs Durante pandemia

```
1 t.test(EJECUTADO ~ Fase, data = subset(data, Fase %in%
      c("Pre-pandemia", "Durante pandemia")))
```

Resultado:

- $t = -16.465$

- p-valor < 2.2e-16
- Media Pre-pandemia = 6038.51 MW
- Media Durante pandemia = 5953.42 MW

Interpretación:

Existe una diferencia estadísticamente significativa. La demanda disminuyó en promedio 85 MW durante la pandemia.

■ Pre-pandemia vs Post-pandemia

```
1 t.test(EJECUTADO ~ Fase, data = subset(data, Fase %in%
  c("Pre-pandemia", "Post-pandemia")))
```

Resultado:

- t = 146.83
- p-valor < 2.2e-16
- Media Pre-pandemia = 6038.51 MW
- Media Post-pandemia = 6681.14 MW

Interpretación:

La demanda post-pandemia aumentó significativamente respecto al periodo pre-pandemia (diferencia de aproximadamente 643 MW).

■ Durante pandemia vs Post-pandemia

```
1 t.test(EJECUTADO ~ Fase, data = subset(data, Fase %in%
  c("Durante pandemia", "Post-pandemia")))
```

Resultado:

- t = -138.94
- p-valor < 2.2e-16
- Media Durante pandemia = 5953.42 MW
- Media Post-pandemia = 6681.14 MW

Interpretación:

La demanda aumentó significativamente en el periodo post-pandemia respecto al periodo pandémico.

Análisis de varianza (ANOVA)

Se aplicó una prueba ANOVA de un factor para comparar simultáneamente las tres etapas.

```
1 modelo_anova <- aov(EJECUTADO ~ Fase, data = data)
2 summary(modelo_anova)
```

Resultado:

- $F = 12980$
- $p\text{-valor} < 2e-16$

Interpretación:

Al menos una de las medias difiere significativamente. Esto justifica realizar un test post hoc.

Prueba post hoc de Tukey (Tukey HSD)

Para identificar cuáles grupos difieren, se aplicó la prueba de comparaciones múltiples de Tukey:

```
1 TukeyHSD(modelo_anova)
```

Figura 4.7

Resultados prueba tukey HSD

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = EJECUTADO ~ Fase, data = data)
##
## $Fase
##                diff          lwr          upr p adj
## Post-pandemia-Durante pandemia 727.7196 716.13915 739.30015 0
## Pre-pandemia-Durante pandemia  85.0877  73.50325  96.67215 0
## Pre-pandemia-Post-pandemia -642.6319 -654.20467 -631.05922 0
```

Fuente: Elaboración propia.

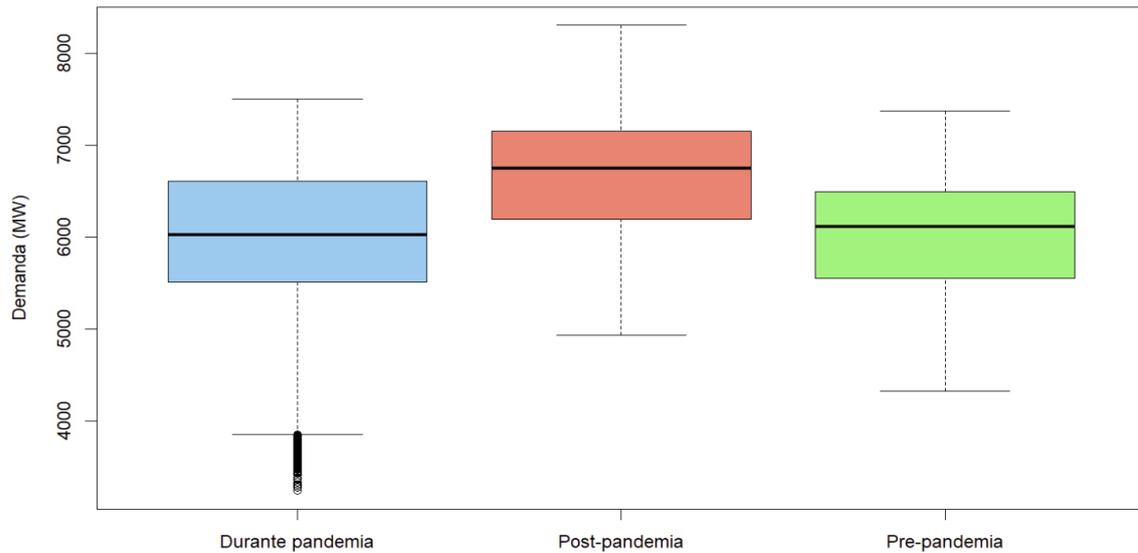
Interpretación:

Todas las diferencias son estadísticamente significativas. Se confirma que la demanda eléctrica cayó durante la pandemia y luego se recuperó superando incluso los niveles pre-pandemia.

Conclusión análisis estadístico inferencial

Los resultados de las pruebas t, ANOVA y Tukey HSD indican que los cambios en la demanda de energía eléctrica entre los tres periodos analizados son estadísticamente significativos. Estos hallazgos respaldan cuantitativamente que la pandemia tuvo un impacto real en los patrones de consumo, cuyos efectos se prolongaron más allá del periodo crítico de confinamiento. La Figura 4.8 confirma esta afirmación, ya que muestra que las distribuciones de demanda en cada fase son claramente diferentes, evidenciado por rangos intercuartílicos distintos.

Figura 4.8
Demanda eléctrica por fase



Fuente: Elaboración propia.

4.3. Normalización de datos

Para asegurar que la serie temporal estuviera en una escala homogénea y evitar que los valores más grandes dominaran el proceso de entrenamiento del modelo, se aplicó una normalización de los datos. Esta técnica es fundamental en el preprocesamiento de series temporales, especialmente cuando se emplean modelos sensibles a la escala, como las redes neuronales LSTM (Hastie et al., 2009).

Se utilizó la función `MinMaxScaler` de la librería `sklearn.preprocessing`, que transforma los datos linealmente para que sus valores se encuentren en un rango definido, en este caso entre 0 y 1. La fórmula aplicada es:

$$X_{\text{scaled}} = \frac{X - \text{mín}(X)}{\text{máx}(X) - \text{mín}(X)} \cdot (\text{max_range} - \text{min_range}) + \text{min_range}$$

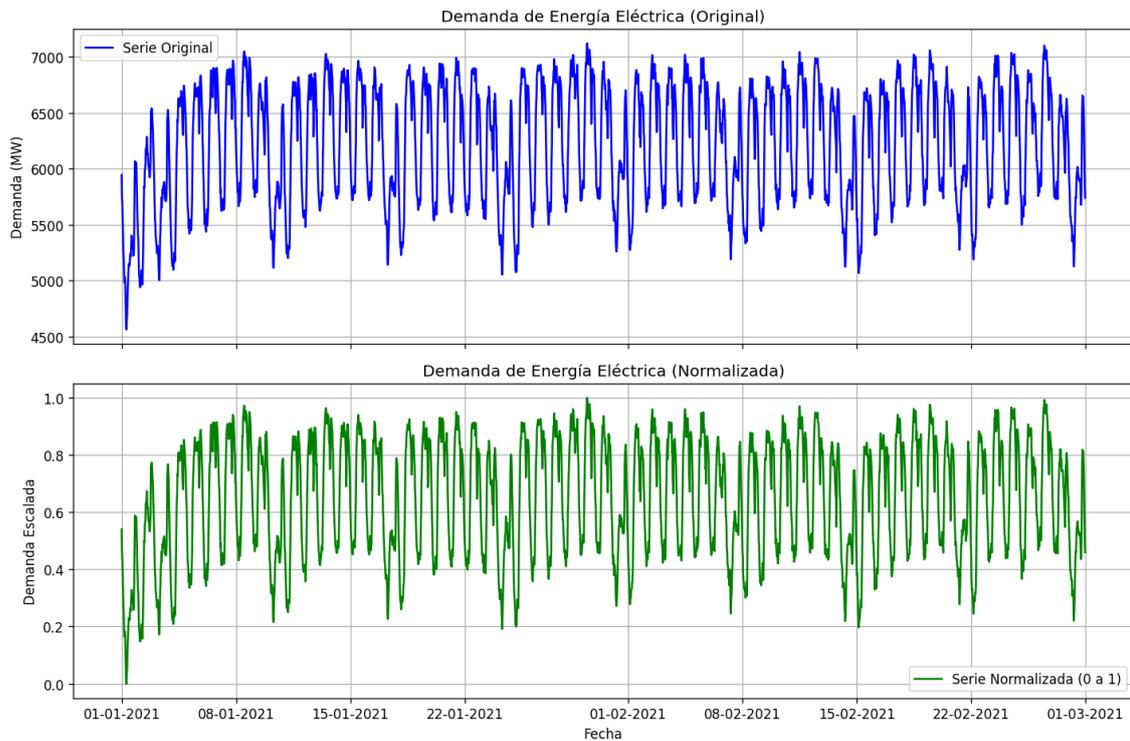
Donde $\text{min}(X)$ y $\text{max}(X)$ son los valores mínimo y máximo de la variable

original, respectivamente.

Se seleccionó el rango de 0 a 1 porque todos los valores de la serie son positivos y esta escala es ampliamente recomendada para mejorar la convergencia y estabilidad de los modelos basados en redes neuronales, especialmente LSTM (Géron, 2019).

En la Figura 4.9 se muestra el resultado del proceso de normalización aplicado a la serie temporal. Es importante destacar que la normalización no afectó el comportamiento general de la serie temporal, preservando tanto las tendencias como las características inherentes de los datos.

Figura 4.9
Normalización de la serie temporal



Fuente: Elaboración propia.

4.4. Definición y entrenamiento

4.4.1. Definición del modelo LSTM seleccionado

Dado que se trabajó con una serie temporal, fue necesario configurar la red neuronal LSTM como un modelo univariado y unistep, lo que implicó emplear una ventana deslizante de valores pasados para predecir un único valor futuro.

4.4.2. Creación de la base de datos supervisada

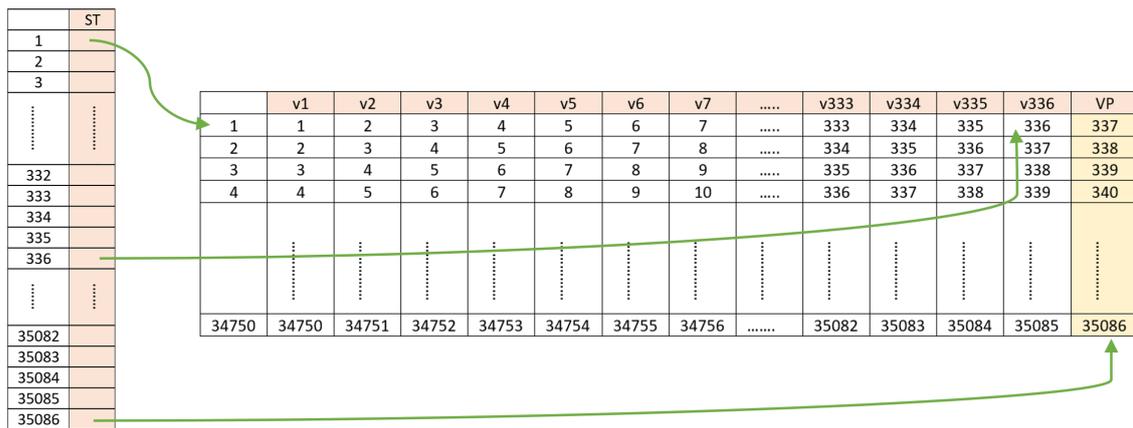
En este estudio se adoptó un enfoque de predicción univariada - unistep empleando redes LSTM, dado que este tipo de modelo, al trabajar únicamente con la variable histórica de interés, ofrece una alta precisión para pronósticos inmediatos con menor complejidad computacional y sin necesidad de incorporar variables exógenas, lo que facilita su implementación y mantenimiento. Tal como señalan Limouni et al. (2022), este enfoque es suficiente y eficiente para anticipar variaciones inmediatas en series temporales energéticas, reduciendo el riesgo de sobreajuste y el tiempo de entrenamiento. Entences se transformo la serie a formato supervisado mediante ventanas deslizantes de longitud fija que capturan el historial reciente, y la definición de la variable objetivo como el valor de la demanda en el siguiente instante temporal. Este procedimiento permitió estructurar el conjunto de entrenamiento y validación de manera coherente con la dinámica secuencial de la demanda eléctrica, optimizando el rendimiento del modelo en la predicción a corto plazo.

Primero se inicializan las variables como arreglos vacíos. A continuación, se ingresa a un bucle que itera sobre los índices del arreglo mientras existan datos suficientes para extraer tanto entradas como salidas. Dentro del bucle, se toman fragmentos de longitud `input_length` para la variable X y se obtienen valores futuros de longitud `output_length` para la variable Y. Posteriormente, estos datos se convierten en arreglos de

NumPy. Finalmente, se devuelven los datos y el proceso concluye.

Con más detalle en la Figura 4.10 se ilustra el proceso de construcción de la tabla supervisada a partir de una serie temporal con comportamiento cíclico semanal. Dado que la serie presenta un ciclo de 336 observaciones (equivalente a una semana), se utilizó una ventana deslizante de ese tamaño como conjunto de variables predictoras para estimar el valor inmediatamente siguiente (el dato 337). Para generar los siguientes ejemplos supervisados, la ventana se desplaza una posición hacia adelante: se toman los datos del 2 al 337 para predecir el dato 338, y así sucesivamente. Este procedimiento se repite de manera continua hasta alcanzar el último valor de la serie (el dato 35,086), permitiendo generar un conjunto de datos estructurado para entrenar el modelo LSTM.

Figura 4.10
Creación base de datos supervisada



Fuente: Elaboración propia.

La Figura 4.11 muestra las primeras muestras del conjunto de datos supervisado generado a partir de la serie de demanda eléctrica preprocesada. Cada muestra contiene 336 valores de entrada (x_{tr}), que corresponden a una semana de registros escalados, y un único valor de salida (y_{tr}), que representa la demanda esperada en la siguiente media hora. Esta estructura es adecuada para entrenar modelos secuenciales como LSTM, ya que permite aprender patrones temporales a partir de ventanas deslizantes. La visualización confirma que los datos fueron correctamente segmentados y escalados, y que la salida (y_{tr}) sigue una tendencia coherente con los últimos valores de entrada, lo que sugiere una preparación correcta para el entrenamiento supervisado.

Figura 4.11

Muestra del conjunto de entrenamiento

```

Muestra 1 (X):
[0.391 0.362 0.327 0.303 0.273 0.266 0.266 0.291 0.296 0.316 0.345 0.401 0.437 0.413 0.456 0.509 0.547 0.616 0.653 0.681 0.699 0.755 0.747 0.737
0.702 0.694 0.676 0.721 0.768 0.757 0.742 0.762 0.763 0.735 0.683 0.676 0.734 0.747 0.75 0.736 0.729 0.713 0.694 0.659 0.619 0.561 0.518 0.465
0.401 0.368 0.327 0.311 0.283 0.287 0.272 0.268 0.285 0.289 0.357 0.403 0.434 0.403 0.467 0.472 0.579 0.604 0.67 0.676 0.728 0.746 0.772 0.752
0.75 0.683 0.705 0.715 0.761 0.707 0.732 0.712 0.732 0.713 0.671 0.647 0.722 0.768 0.758 0.758 0.742 0.727 0.698 0.676 0.631 0.563 0.486 0.447
0.399 0.372 0.352 0.324 0.305 0.297 0.265 0.261 0.305 0.328 0.359 0.393 0.451 0.446 0.485 0.499 0.583 0.656 0.708 0.698 0.768 0.764 0.774 0.764
0.751 0.717 0.721 0.741 0.776 0.752 0.769 0.755 0.744 0.711 0.69 0.681 0.77 0.738 0.76 0.75 0.739 0.731 0.715 0.688 0.647 0.596 0.558 0.492
0.413 0.409 0.382 0.353 0.336 0.316 0.31 0.295 0.311 0.335 0.315 0.343 0.336 0.317 0.383 0.444 0.545 0.57 0.65 0.666 0.67 0.695 0.735 0.73
0.702 0.698 0.648 0.623 0.632 0.644 0.629 0.586 0.595 0.615 0.602 0.627 0.725 0.801 0.772 0.74 0.789 0.746 0.714 0.731 0.629 0.548 0.498 0.418
0.387 0.384 0.343 0.318 0.251 0.282 0.251 0.246 0.227 0.238 0.271 0.251 0.17 0.135 0.148 0.209 0.256 0.278 0.292 0.308 0.361 0.4 0.407 0.382
0.404 0.417 0.386 0.393 0.405 0.402 0.401 0.408 0.406 0.425 0.445 0.477 0.62 0.68 0.683 0.69 0.697 0.677 0.65 0.603 0.549 0.466 0.398 0.35
0.321 0.279 0.251 0.249 0.211 0.204 0.184 0.179 0.194 0.237 0.283 0.331 0.407 0.41 0.412 0.467 0.552 0.587 0.635 0.675 0.728 0.727 0.743 0.729
0.745 0.717 0.695 0.719 0.741 0.744 0.73 0.718 0.726 0.701 0.699 0.645 0.725 0.764 0.769 0.762 0.761 0.75 0.718 0.709 0.663 0.588 0.516 0.473
0.402 0.381 0.349 0.314 0.303 0.286 0.28 0.307 0.307 0.344 0.347 0.389 0.445 0.453 0.484 0.521 0.581 0.61 0.63 0.67 0.693 0.702 0.714 0.69
0.657 0.627 0.614 0.641 0.69 0.686 0.683 0.68 0.684 0.683 0.689 0.639 0.717 0.756 0.764 0.75 0.738 0.73 0.694 0.675 0.641 0.562 0.506 0.458]
Target (y_tr[0]) como vector columna:
[[0.405]]
-----
Muestra 2 (X):
[0.362 0.327 0.303 0.273 0.266 0.266 0.291 0.296 0.316 0.345 0.401 0.437 0.413 0.456 0.509 0.547 0.616 0.653 0.681 0.699 0.755 0.747 0.737 0.702
0.694 0.676 0.721 0.768 0.757 0.742 0.762 0.763 0.735 0.683 0.676 0.734 0.747 0.75 0.736 0.729 0.713 0.694 0.659 0.619 0.561 0.518 0.465 0.401
0.368 0.327 0.311 0.283 0.287 0.272 0.268 0.285 0.289 0.357 0.403 0.434 0.403 0.467 0.472 0.579 0.604 0.67 0.676 0.728 0.746 0.772 0.752 0.75
0.683 0.705 0.715 0.761 0.707 0.732 0.712 0.732 0.713 0.671 0.647 0.722 0.768 0.758 0.758 0.742 0.727 0.698 0.676 0.631 0.563 0.486 0.447 0.399
0.372 0.352 0.324 0.305 0.297 0.265 0.261 0.305 0.328 0.359 0.393 0.451 0.446 0.485 0.499 0.583 0.656 0.708 0.698 0.768 0.764 0.774 0.764 0.751
0.717 0.721 0.741 0.776 0.752 0.769 0.755 0.744 0.711 0.69 0.681 0.77 0.738 0.76 0.75 0.739 0.731 0.715 0.688 0.647 0.596 0.558 0.492 0.413
0.409 0.382 0.353 0.336 0.316 0.31 0.295 0.311 0.335 0.315 0.343 0.336 0.317 0.383 0.444 0.545 0.57 0.65 0.666 0.67 0.695 0.735 0.73 0.702
0.698 0.648 0.623 0.632 0.644 0.629 0.586 0.595 0.615 0.602 0.627 0.725 0.801 0.772 0.74 0.789 0.746 0.714 0.731 0.629 0.548 0.498 0.418 0.387
0.384 0.343 0.318 0.251 0.282 0.251 0.246 0.227 0.238 0.271 0.251 0.17 0.135 0.148 0.209 0.256 0.278 0.292 0.308 0.361 0.4 0.407 0.382 0.404
0.417 0.386 0.393 0.405 0.402 0.401 0.408 0.406 0.425 0.445 0.477 0.62 0.68 0.683 0.69 0.697 0.677 0.65 0.603 0.549 0.466 0.398 0.35 0.321
0.279 0.251 0.249 0.211 0.204 0.184 0.179 0.194 0.237 0.283 0.331 0.407 0.41 0.412 0.467 0.552 0.587 0.635 0.675 0.728 0.727 0.743 0.729 0.745
0.717 0.695 0.719 0.741 0.744 0.73 0.718 0.726 0.701 0.699 0.645 0.725 0.764 0.769 0.762 0.761 0.75 0.718 0.709 0.663 0.588 0.516 0.473 0.402
0.381 0.349 0.314 0.303 0.286 0.28 0.307 0.307 0.344 0.347 0.389 0.445 0.453 0.484 0.521 0.581 0.61 0.63 0.67 0.693 0.702 0.714 0.69 0.657
0.627 0.614 0.641 0.69 0.686 0.683 0.68 0.684 0.683 0.689 0.639 0.717 0.756 0.764 0.75 0.738 0.73 0.694 0.675 0.641 0.562 0.506 0.458 0.405]
Target (y_tr[1]) como vector columna:
[[0.358]]
-----
Muestra 3 (X):
[0.327 0.303 0.273 0.266 0.266 0.291 0.296 0.316 0.345 0.401 0.437 0.413 0.456 0.509 0.547 0.616 0.653 0.681 0.699 0.755 0.747 0.737 0.702 0.694
0.676 0.721 0.768 0.757 0.742 0.762 0.763 0.735 0.683 0.676 0.734 0.747 0.75 0.736 0.729 0.713 0.694 0.659 0.619 0.561 0.518 0.465 0.401 0.368
0.327 0.311 0.283 0.287 0.272 0.268 0.285 0.289 0.357 0.403 0.434 0.403 0.467 0.472 0.579 0.604 0.67 0.676 0.728 0.746 0.772 0.752 0.75 0.683
0.705 0.715 0.761 0.707 0.732 0.712 0.732 0.713 0.671 0.647 0.722 0.768 0.758 0.758 0.742 0.727 0.698 0.676 0.631 0.563 0.486 0.447 0.399 0.372
0.352 0.324 0.305 0.297 0.265 0.261 0.305 0.328 0.359 0.393 0.451 0.446 0.485 0.499 0.583 0.656 0.708 0.698 0.768 0.764 0.774 0.764 0.751 0.717
0.721 0.741 0.776 0.752 0.769 0.755 0.744 0.711 0.69 0.681 0.77 0.738 0.76 0.75 0.739 0.731 0.715 0.688 0.647 0.596 0.558 0.492 0.413 0.409
0.382 0.353 0.336 0.316 0.31 0.295 0.311 0.335 0.315 0.343 0.336 0.317 0.383 0.444 0.545 0.57 0.65 0.666 0.67 0.695 0.735 0.73 0.702 0.698
0.648 0.623 0.632 0.644 0.629 0.586 0.595 0.615 0.602 0.627 0.725 0.801 0.772 0.74 0.789 0.746 0.714 0.731 0.629 0.548 0.498 0.418 0.387 0.384
0.343 0.318 0.251 0.282 0.251 0.246 0.227 0.238 0.271 0.251 0.17 0.135 0.148 0.209 0.256 0.278 0.292 0.308 0.361 0.4 0.407 0.382 0.404 0.417
0.386 0.393 0.405 0.402 0.401 0.408 0.406 0.425 0.445 0.477 0.62 0.68 0.683 0.69 0.697 0.677 0.65 0.603 0.549 0.466 0.398 0.35 0.321 0.279
0.251 0.249 0.211 0.204 0.184 0.179 0.194 0.237 0.283 0.331 0.407 0.41 0.412 0.467 0.552 0.587 0.635 0.675 0.728 0.727 0.743 0.729 0.745 0.717
0.695 0.719 0.741 0.744 0.73 0.718 0.726 0.701 0.699 0.645 0.725 0.764 0.769 0.762 0.761 0.75 0.718 0.709 0.663 0.588 0.516 0.473 0.402 0.381
0.349 0.314 0.303 0.286 0.28 0.307 0.307 0.344 0.347 0.389 0.445 0.453 0.484 0.521 0.581 0.61 0.63 0.67 0.693 0.702 0.714 0.69 0.657 0.627
0.614 0.641 0.69 0.686 0.683 0.68 0.684 0.683 0.689 0.639 0.717 0.756 0.764 0.75 0.738 0.73 0.694 0.675 0.641 0.562 0.506 0.458 0.405 0.358]
Target (y_tr[2]) como vector columna:
[[0.337]]
-----
Muestra 4 (X):
[0.303 0.273 0.266 0.266 0.291 0.296 0.316 0.345 0.401 0.437 0.413 0.456 0.509 0.547 0.616 0.653 0.681 0.699 0.755 0.747 0.737 0.702 0.694 0.676
0.721 0.768 0.757 0.742 0.762 0.763 0.735 0.683 0.676 0.734 0.747 0.75 0.736 0.729 0.713 0.694 0.659 0.619 0.561 0.518 0.465 0.401 0.368 0.327

```

Fuente: Elaboración propia.

4.4.3. Justificación del modelo seleccionado

El modelo seleccionado para esta investigación fue una red neuronal Long Short-Term Memory (LSTM), conocida por su capacidad de capturar dependencias a largo plazo en series temporales. Este modelo se consideró ideal para abordar el problema, dado que podía identificar patrones complejos y fluctuaciones significativas en la demanda de energía eléctrica, particularmente durante la pandemia de COVID-19.

Se configuró una única capa LSTM con 256 unidades, proporcionando un equilibrio entre la capacidad de representación y el costo computacional. Además, se añadió una capa densa con activación lineal para predecir un único paso de salida, lo que permitió mantener la granularidad de los datos y realizar predicciones a 30 minutos hacia adelante.

Para entrenar el modelo se emplearon datos de la serie temporal correspondientes a la etapa pre-pandemia. Estos se dividieron en un 80 % para entrenamiento, un 10 % para prueba (test) y un 10 % para validación.

4.4.4. Configuración de hiperparámetros

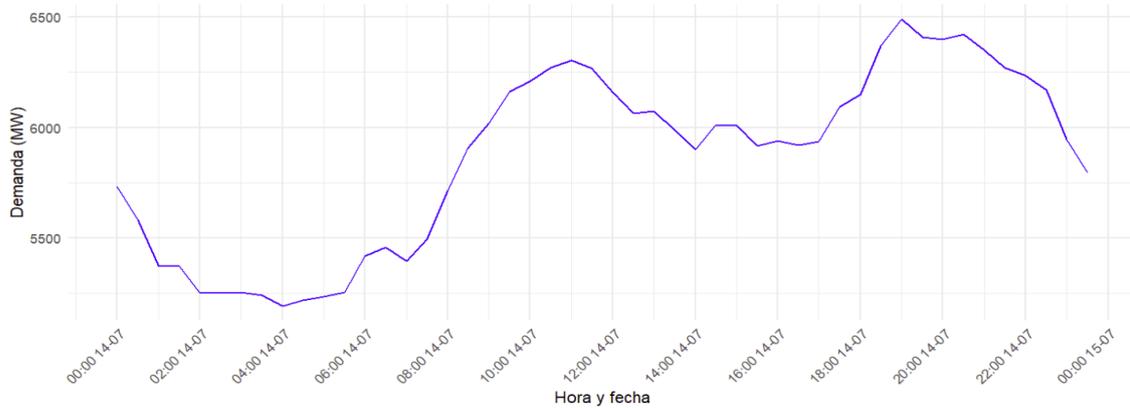
La serie temporal de demanda de energía eléctrica presenta varios ciclos identificados. Cabe recordar que, por día, se registran 48 datos. Entre los ciclos detectados se encuentran:

- **Ciclo por día:** Cada día exhibe un comportamiento similar. Durante las primeras horas de la madrugada, la demanda disminuye, alcanzando sus niveles más bajos entre las 2:00 a.m. y las 6:00 a.m. A partir de las 6:30 a.m., la demanda comienza a aumentar, alcanzando su pico más alto alrededor de las 11:00 a.m. Luego desciende hasta aproximadamente las 4:30 p.m., para posteriormente volver a subir, logrando su máximo diario entre las 6:30 p.m. y las 9:00 p.m.

Después de ese horario, la demanda vuelve a disminuir. Este patrón se aprecia en la Figura 4.12, donde en el eje X se puede ver en el formato hora y fecha.

Figura 4.12

Comportamiento demanda de energía eléctrica por día

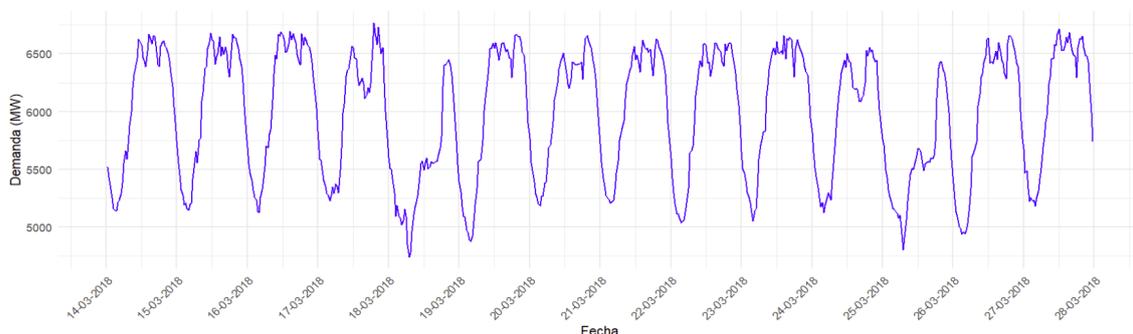


Fuente: Elaboración propia.

- **Ciclo semanal:** Se observó un patrón semanal donde cada día de la semana tiene un patrón singular. Por ejemplo, sábados y domingos presentan un comportamiento distinto al de los demás días. Además, los viernes suelen mostrar picos superiores con respecto al resto de la semana. Este comportamiento se ilustra en la Figura 4.13, donde se puede visualizar 2 semanas desde el 14 de marzo al 28 de marzo del 2018, claramente se puede observar que los días 18, 19, 25 y 26 son días sábado y domingo.

Figura 4.13

Comportamiento demanda de energía eléctrica por semana



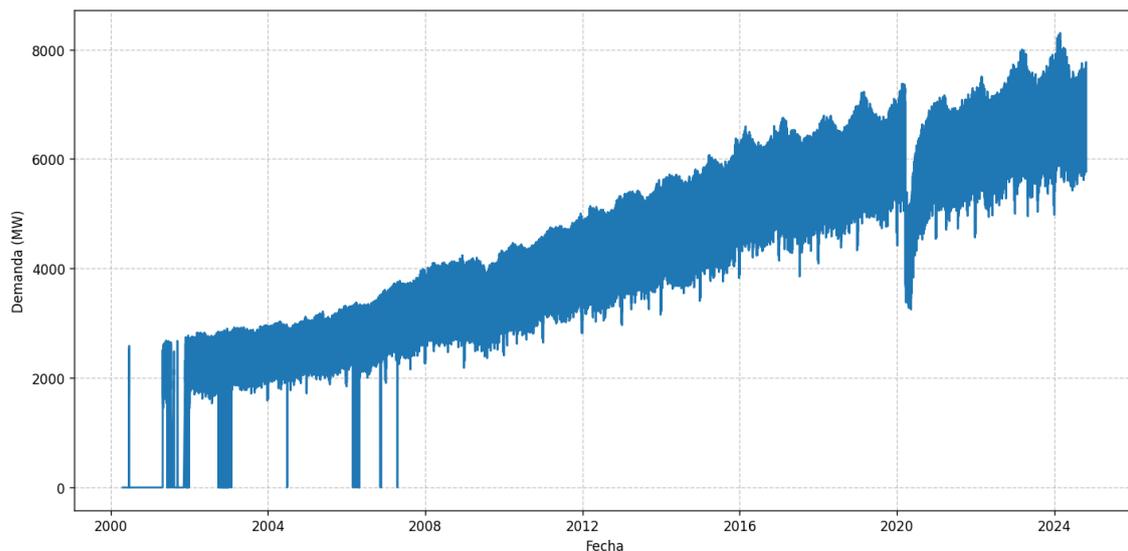
Fuente: Elaboración propia.

- **Ciclo anual:** La diversa geografía del Perú determina condiciones climáticas marcadas, como períodos de lluvias y estiaje. Asimismo, su ubicación al sur de

la línea ecuatorial implica ciclos en la duración del día y la noche. Estos factores influyen en la demanda de energía, generando un ciclo anual observable, tal como se muestra en la Figura 4.14, se puede observar que hasta el 2008 se tiene datos faltantes, a partir del 2010 se aprecia picos anuales generalmente en los meses de diciembre a abril.

Figura 4.14

Comportamiento demanda de energía eléctrica por año



Fuente: Elaboración propia.

Dada la existencia de estos ciclos, se desarrollaron varios modelos modificando la longitud de entrada (`input_length`) con valores de 24, 48, 336, 1440 y 17,520. Entre todos ellos, el modelo con longitud de entrada correspondiente al ciclo semanal presentó la menor pérdida.

El resto de los hiperparámetros del modelo, como el número de unidades en la capa LSTM, la tasa de aprendizaje, y el tamaño del batch, se definieron considerando las características específicas de la serie temporal y siguiendo las mejores prácticas documentadas en la literatura especializada sobre redes LSTM para series temporales (Géron (2019); Brownlee (2018)). Estas prácticas incluyen ajustar la arquitectura del modelo para capturar patrones temporales relevantes, evitar el sobreajuste mediante técnicas de regularización, y seleccionar parámetros de entrenamiento que optimicen la convergencia y la generalización.

En la Tabla 4.3 se detallan los valores asignados y su justificación.

Tabla 4.3

Hiperparámetros seleccionados y sus justificaciones.

Hiperparámetro	Valor	Justificación
Longitud de entrada	336 pasos	Corresponde a una semana de datos (48 pasos diarios por 7 días), capturando tendencias semanales.
Longitud de salida	1 paso	Predicción inmediata a 30 minutos, ideal para mantener la granularidad de los datos.
Número de unidades LSTM	256	Suficiente para capturar patrones complejos sin sobrecargar el modelo.
Tasa de aprendizaje	0.00005	Baja para permitir ajustes precisos en cada iteración.
Tamaño del batch	64	Balance entre eficiencia computacional y capacidad de generalización.
Épocas	50	Adecuado para alcanzar la convergencia sin riesgo de sobreentrenamiento.

```
1 # Preparar datos para entrenamiento y prediccion
2 INPUT_LENGTH = 336 # 1 semana de datos
3 OUTPUT_LENGTH = 1 # la siguiente media hora
4
5 # Escalar datos pre-pandemia
6 pre_pandemia_scaled, scaler = escalar_dataset(pre_pandemia.
7     values)
8
9 # Crear datasets supervisados para entrenamiento
10 x_tr, y_tr = crear_dataset_supervisado(pre_pandemia_scaled,
11     INPUT_LENGTH, OUTPUT_LENGTH)
12
13 modelo = Sequential()
14 modelo.add(LSTM(N_UNITS, input_shape=INPUT_SHAPE))
15 modelo.add(Dense(OUTPUT_LENGTH, activation='linear'))
```

```
16
17 modelo.compile(
18     optimizer=RMSprop(learning_rate=5e-5),
19     loss=root_mean_squared_error)
20 modelo.summary()
21 EPOCHS = 50
22 BATCH_SIZE = 64
23 historia = modelo.fit(
24     x_tr,
25     y_tr,
26     epochs=EPOCHS,
27     batch_size=BATCH_SIZE,
28     validation_data = (x_tr, y_tr),
29     verbose=2)
```

4.4.5. Entrenamiento del modelo

Los datos correspondientes al periodo pre-pandemia fueron escalados utilizando el método Min-Max, asegurando que se encontraran en el rango $[0,1]$. Posteriormente, estos datos se estructuraron en un formato supervisado, donde 336 pasos de entrada se utilizaron para predecir un único paso de salida.

El modelo fue entrenado utilizando el optimizador RMSprop, seleccionado por su eficacia en problemas secuenciales, y una función de pérdida personalizada basada en la raíz del error cuadrático medio (RMSE), adecuada para evaluar la precisión de las predicciones continuas. El entrenamiento se realizó durante 50 épocas, con un tamaño de batch de 64, y los resultados mostraron una disminución continua de la función de pérdida, alcanzando un valor de 0.0369.

4.4.6. Evaluación del modelo

El modelo fue evaluado utilizando el RMSE, una métrica adecuada para medir la discrepancia entre las predicciones y los valores reales en la misma escala. Los resultados durante la pandemia y el periodo post-pandemia fueron los siguientes:

- RMSE durante la pandemia: 0.0304.
- RMSE post-pandemia: 0.0295

Estos valores indicaron que el modelo mantuvo una buena capacidad predictiva en ambas fases. El RMSE ligeramente menor en el periodo post-pandemia sugirió que las condiciones de consumo más estables durante la recuperación permitieron al modelo generalizar mejor.

4.4.7. Interpretación de resultados del modelo de predicción

El modelo LSTM fue capaz de capturar correctamente las tendencias principales de la serie temporal. La baja tasa de aprendizaje utilizada permitió ajustes graduales en los pesos, lo que resultó en un entrenamiento estable y una convergencia adecuada. La precisión del modelo fue consistente tanto en el periodo de la pandemia como en el periodo post-pandemia, con una ligera mejora en este último.

Los resultados obtenidos validaron la eficacia del modelo para predecir la demanda de energía eléctrica, incluso en escenarios complejos y disruptivos como la pandemia de COVID-19. Esto sentó una base sólida para su aplicación en análisis más profundos y en la identificación de anomalías.

4.5. Detección de anomalías

En esta investigación, el término tipo de anomalías se refiere a la naturaleza del comportamiento inusual detectado en la serie temporal, tales como picos atípicos, caídas abruptas o alteraciones en la tendencia y estacionalidad. La magnitud alude a la desviación cuantitativa (en megavatios, MW) respecto al valor esperado que predice el modelo LSTM entrenado con datos del periodo pre-pandemia.

4.5.1. Metodología aplicada

En este punto, se utilizó el modelo entrenado para realizar las predicciones en las etapas durante y después de la pandemia, abarcando el periodo del 14 de marzo de 2020 al 13 de marzo de 2024, respectivamente.

Las predicciones se reescalaron a su rango original utilizando el mismo escalador empleado durante la normalización.

Finalmente, se calcularon los errores como la diferencia entre los valores predichos y los valores reales en cada caso.

$$error_i = real_i - predicho_i$$

4.5.2. Umbral para identificar anomalías

Para identificar valores atípicos en la serie temporal durante el periodo de la pandemia, se definió un umbral basado en el percentil 95 de los errores absolutos entre las predicciones del modelo y los valores observados. Esta elección permite establecer un límite superior que captura la mayoría de los errores normales, considerando como anomalías aquellas desviaciones que superan este umbral.

El uso del percentil 95 es una práctica común en la detección de anomalías, ya que se enfoca en detectar eventos raros o extremos que se encuentran en el 5 % superior de la distribución de errores, evitando al mismo tiempo la detección excesiva de falsos positivos. Además, este umbral es robusto ante la presencia de valores atípicos en la muestra, y se adapta a la distribución empírica de los errores (Chandola et al., 2009; Ahmed et al., 2016).

La fórmula aplicada es:

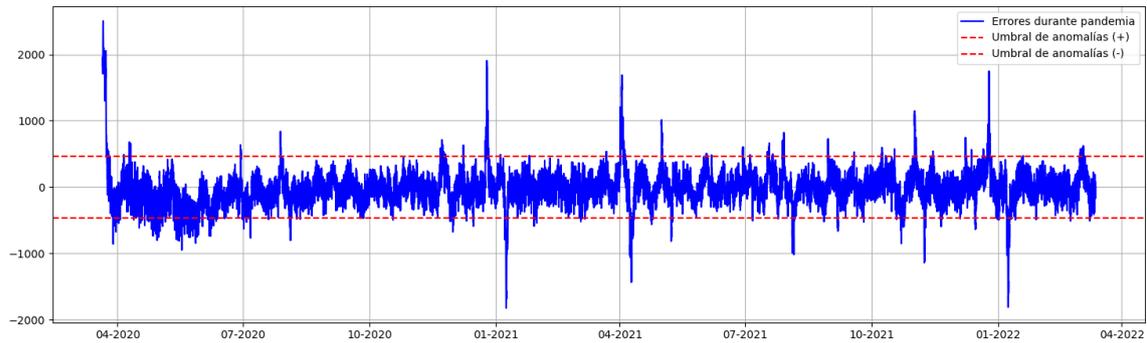
$$umbral = percentil_{95}(|errores_durante_la_pandemia|)$$

Por lo tanto, se consideran anomalías aquellos puntos donde el valor absoluto del error excede este umbral definido..

4.5.3. Cantidad de anomalías detectadas

- La Figura 4.15 presenta la serie temporal de los errores del modelo de predicción de demanda de energía eléctrica durante la pandemia, donde la línea azul representa los errores y las dos líneas rojas discontinuas definen el Umbral de anomalías, marcando el rango de errores esperados. Los puntos de la serie de errores que exceden estos umbrales rojos son clasificados como anomalías, indicando desviaciones significativas del modelo. Como se observa, el modelo identificó un considerable número de 1,733 anomalías durante este periodo, un hecho esperable dado los cambios abruptos e impredecibles en la demanda de energía eléctrica causados por las medidas de aislamiento y otras alteraciones en los patrones de consumo durante la pandemia.

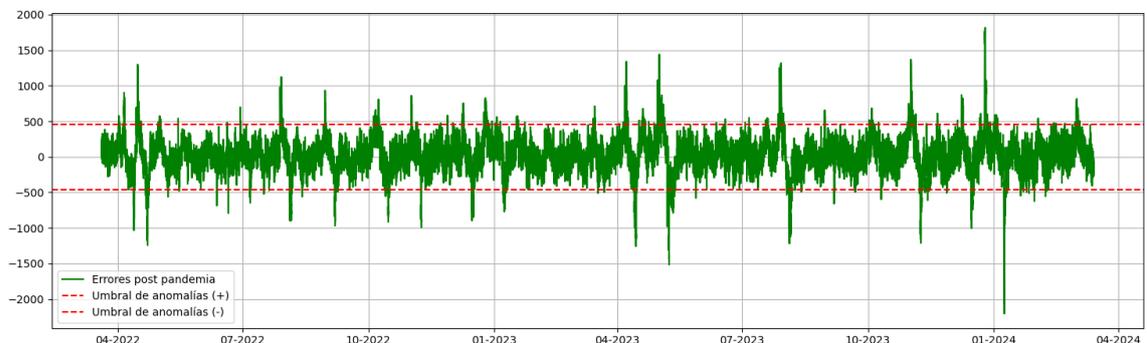
Figura 4.15
Anomalías durante la pandemia



Fuente: Elaboración propia.

- La Figura 4.16 presenta la serie temporal de los errores del modelo de predicción de demanda de energía eléctrica en el periodo post-pandemia, con la línea verde mostrando los errores y las líneas rojas discontinuas marcando el Umbral de anomalías que define el rango de errores esperados; cualquier punto que supere estos umbrales rojos es clasificado como una anomalía, señalando una desviación significativa. En este periodo, se detectaron 1,916 anomalías, lo que, a pesar de la estabilización general, sugiere la persistencia de eventos atípicos posiblemente vinculados a factores externos o a los continuos ajustes en los patrones de consumo post-recuperación. Pero también cabe indicar que el Peru en los años de post-pandemia esta viviendo problemas políticos sociales.

Figura 4.16
Anomalías post-pandemia



Fuente: Elaboración propia.

Al comparar ambos periodos se registraron un número elevado de anomalías que exceden los umbrales rojos definidos para los errores del modelo (1,733 durante

la pandemia y 1,916 en el periodo post-pandemia), la naturaleza y las implicaciones de estas anomalías difieren. Durante la pandemia, el alto número de anomalías era predecible debido a los cambios drásticos y sin precedentes en la demanda de energía causados por las restricciones y el aislamiento; sin embargo, el incremento de anomalías en el periodo post-pandemia, a pesar de una aparente estabilización del contexto general, sugiere que la nueva normalidad y problemas sociales ha introducido variaciones persistentes o nuevas dinámicas en el consumo que el modelo aún no logra capturar plenamente, indicando la necesidad de una adaptación continua del modelo a estos patrones evolutivos de la demanda. Aunque en la etapa post-pandemia se identificó un mayor número de anomalías, la magnitud de estas se redujo considerablemente en comparación con el periodo de pandemia, lo que sugiere una progresiva estabilización en los patrones de consumo energético.

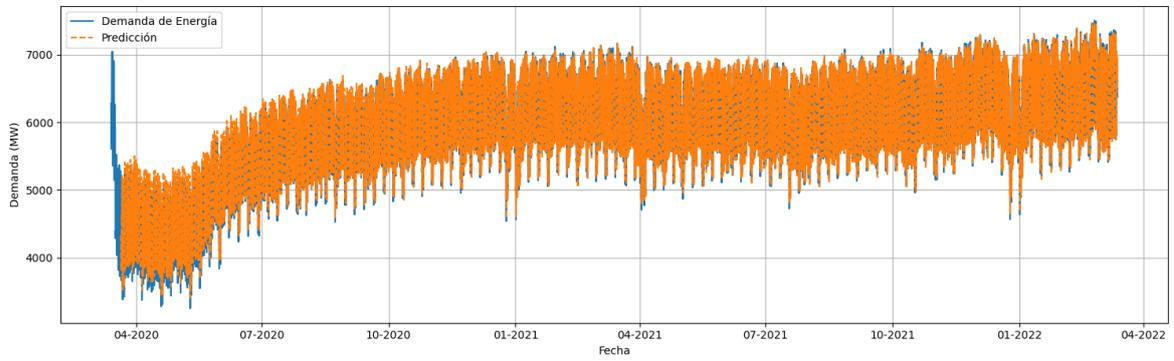
4.5.4. Visualización de anomalías

Las imágenes proporcionan una visión general del comportamiento de la demanda eléctrica durante la pandemia y el rendimiento del modelo LSTM. En la Figura 4.17a muestra la comparación entre la demanda real y las predicciones del modelo durante todo el periodo de la pandemia. Aunque las predicciones (línea naranja) siguen de cerca la serie real (línea azul), se observan pequeñas discrepancias en ciertos picos de alta y baja demanda. En la Figura 4.17b detalla el comportamiento de la serie temporal durante 4 días específicos. Era de esperarse que el modelo capturan parcialmente las oscilaciones diarias de la demanda, con ligeras desviaciones en los picos y en la Figura 4.17c presenta los errores entre la demanda real y las predicciones del modelo, con los valores que exceden un umbral marcado como anomalías (puntos naranjas). Este análisis revela la detección de eventos atípicos en el periodo pandémico, que podrían estar relacionados con cambios drásticos en el consumo eléctrico. En conjunto, los gráficos destacan un modelo que predice bien en términos generales, pero también identifica desviaciones significativas que pueden ser consideradas anomalías importantes para

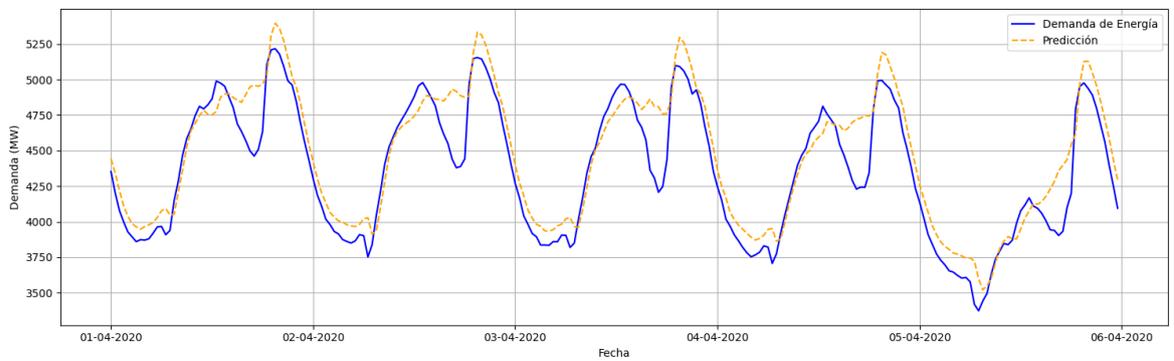
análisis más detallados.

Figura 4.17

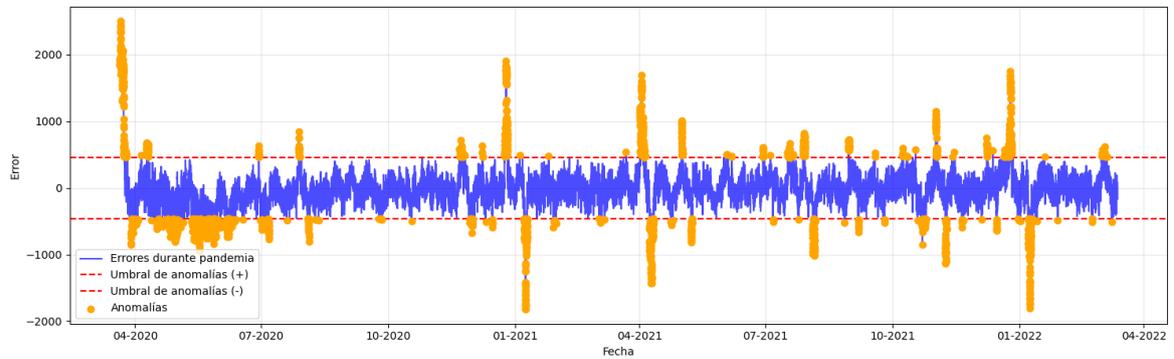
Comparación de datos reales y predicciones durante la pandemia.



(a)



(b)



(c)

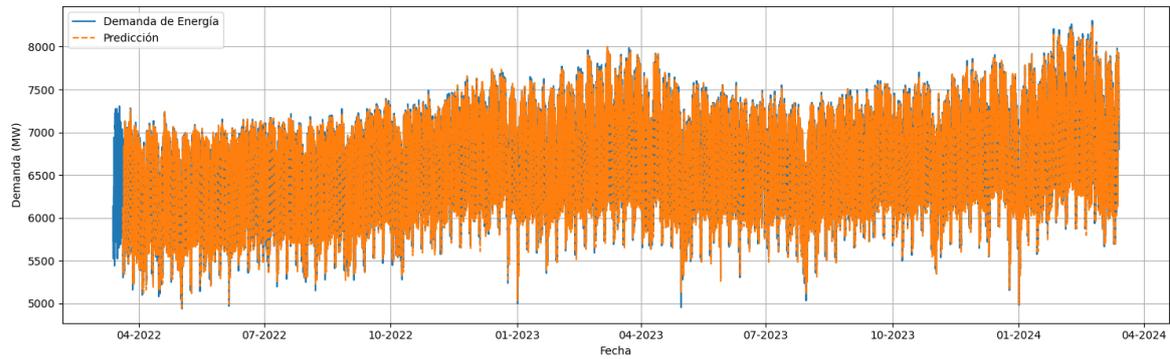
Fuente: Elaboración propia.

4.6. Análisis post pandemia

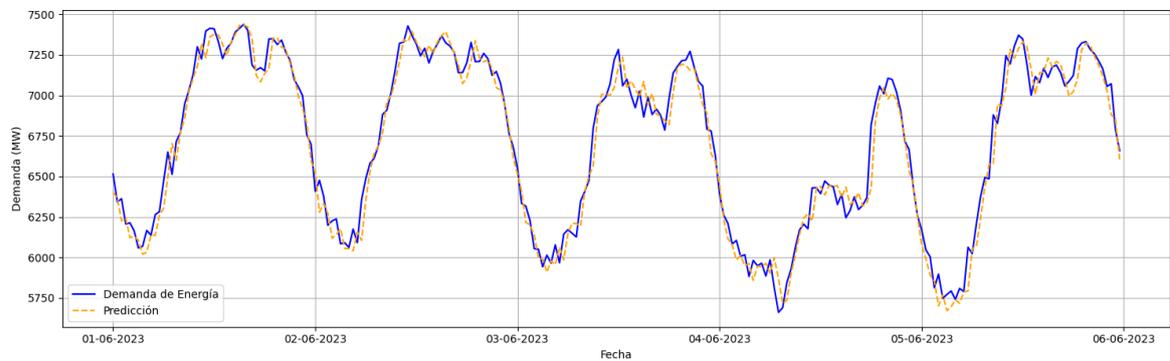
Las imágenes a continuación proporcionan un análisis detallado de la demanda de energía eléctrica y las predicciones post-pandemia utilizando el modelo LSTM: la Figura 4.18a representa la serie completa de la demanda eléctrica post-pandemia (línea azul) comparada con las predicciones del modelo (línea naranja). El modelo sigue de manera precisa las fluctuaciones de la demanda, aunque se observan pequeñas desviaciones en los picos máximos y mínimos, la Figura 4.18b detalla el comportamiento de la demanda y las predicciones para un rango de 4 días específicos. Las predicciones capturan con precisión las oscilaciones diarias, reflejando los patrones normales, aunque hay ligeros desfases en algunos puntos y la Figura 4.18c muestra los errores entre las predicciones y los valores reales, con las anomalías (valores fuera del umbral definido) marcadas en naranja. Esto evidencia eventos fuera de lo esperado en la serie, que podrían reflejar situaciones atípicas post-pandemia. Por lo general el modelo tiene un buen desempeño al capturar los patrones normales post-pandemia, identificando de manera efectiva las anomalías en la serie temporal.

Figura 4.18

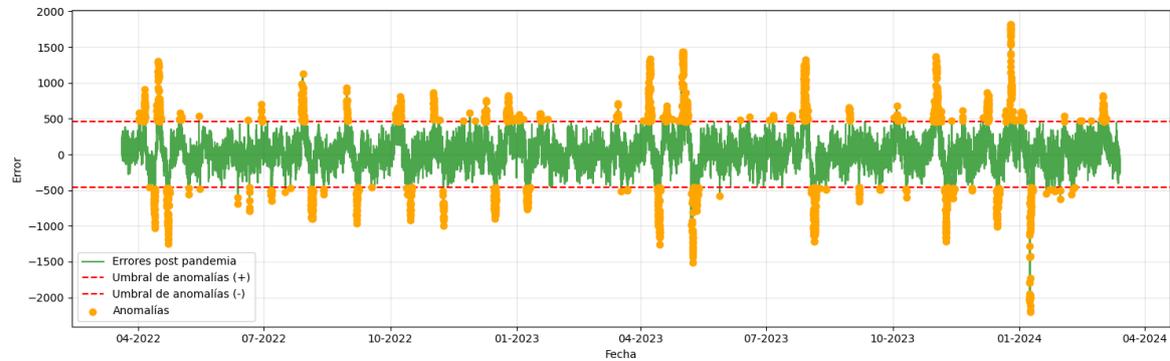
Comparación de datos reales y predicciones post-pandemia



(a)



(b)



(c)

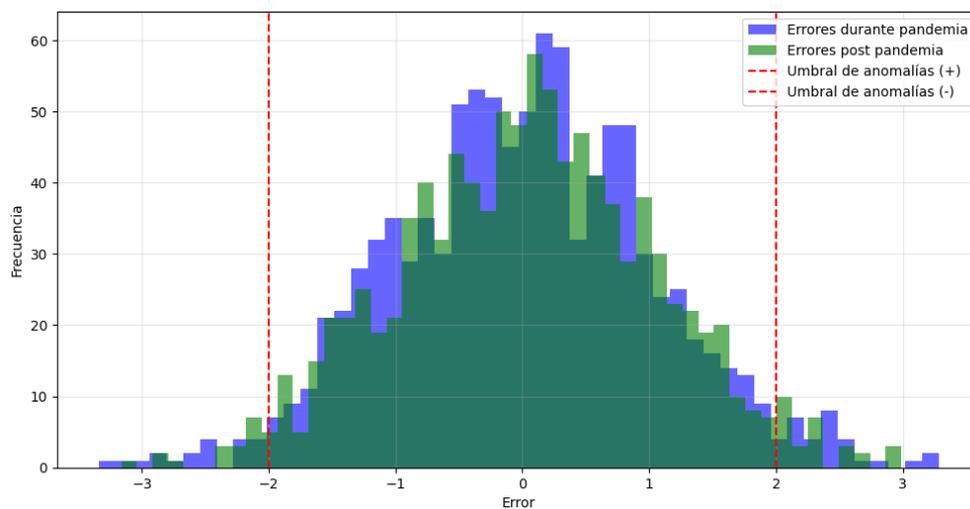
Fuente: Elaboración propia.

La Figura 4.19 compara la distribución de errores entre el periodo de la pandemia y post-pandemia. Ambas distribuciones presentan una forma aproximadamente simétrica y centrada cerca de cero, lo que indica que el modelo predice razonablemente bien en ambos periodos para la mayoría de los datos. Sin embargo, los errores más altos (positivos y negativos) están fuera del umbral marcado, clasificándose como anomalías.

Durante la pandemia los errores parecen estar más dispersos y ligeramente más sesgados hacia valores extremos, lo que sugiere que el modelo tuvo más dificultades para capturar los patrones en este periodo, posiblemente debido a cambios inusuales en la demanda y durante post-pandemia La distribución de errores es más estrecha y concentrada alrededor de cero, lo que indica un mejor ajuste del modelo en este periodo. Las líneas rojas muestran los umbrales de ± 2 en el error. Se observa una mayor frecuencia de errores que exceden estos límites durante la pandemia, lo que evidencia una mayor cantidad de anomalías. Esto podría deberse a la volatilidad y cambios abruptos en los patrones de consumo eléctrico durante ese periodo. En conclusión el histograma revela que los errores post-pandemia son más concentrados alrededor de cero en comparación con los de la pandemia, lo que indica que el modelo predice mejor en este periodo. Aunque las anomalías persisten post-pandemia, su frecuencia es notablemente menor, sugiriendo que los patrones anómalos observados durante la pandemia no se mantienen con la misma intensidad. Esto evidencia una estabilización de la demanda eléctrica en el periodo post-pandemia.

Figura 4.19

Histograma de distribución de errores absolutos durante y post-pandemia

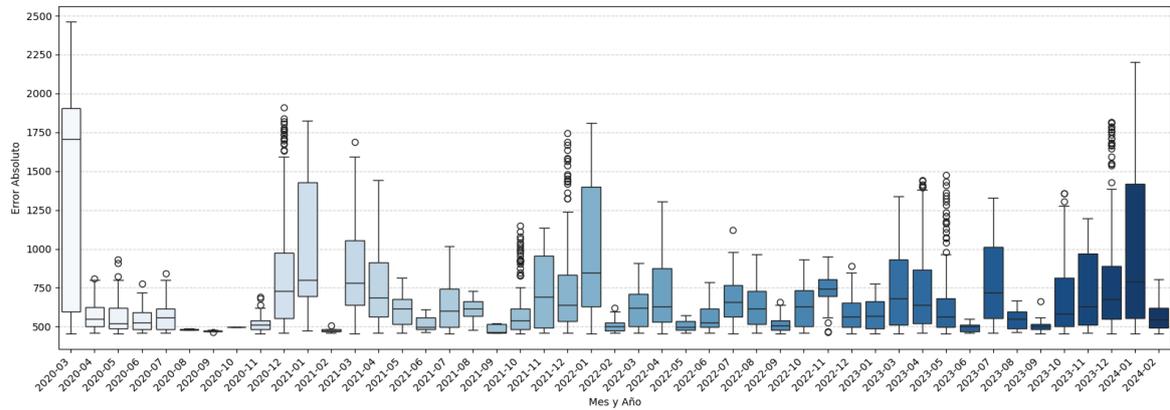


Fuente: Elaboración propia.

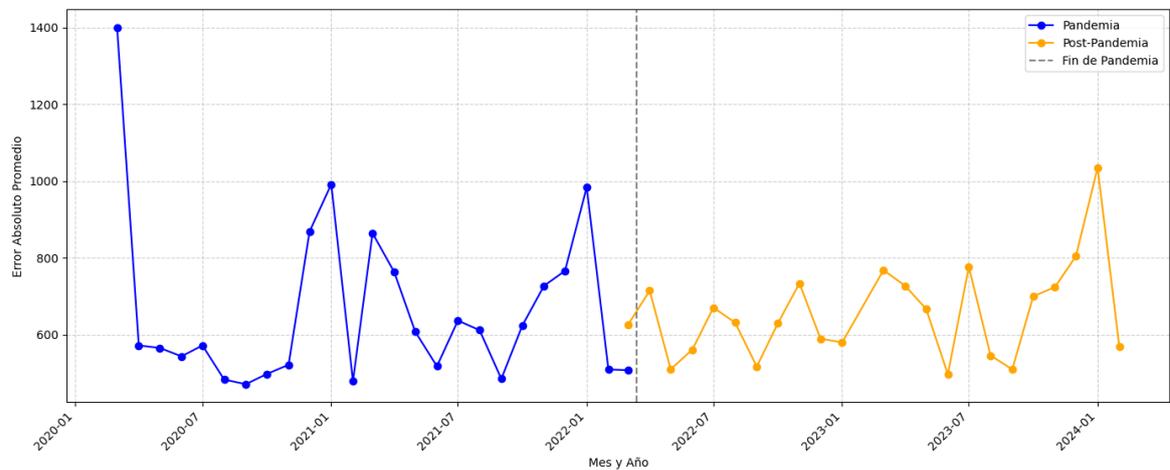
La Figura 4.20 presenta un análisis temporal de las anomalías detectadas en la demanda de energía eléctrica durante los periodos de pandemia y post-pandemia. En la subfigura (a), se muestra la distribución mensual de la amplitud del error absoluto mediante diagramas de caja (boxplots). Se observa una alta dispersión de los errores durante la pandemia, especialmente en los primeros meses de 2020 y en diciembre de 2021, donde se registran valores atípicos numerosos y amplitudes que superan los 2000 MW. A lo largo de los años, las distribuciones tienden a estabilizarse, aunque con repuntes ocasionales hacia finales del periodo post-pandemia, como en diciembre de 2023 y febrero de 2024. Por su parte, la subfigura (b) presenta el promedio mensual del error absoluto. Durante la pandemia (en azul), el error promedio mostró una alta volatilidad, con un pico destacado en marzo de 2020 que alcanzó los 1400 MW, seguido de oscilaciones marcadas hasta inicios de 2022. En la etapa post-pandemia (en naranja), el promedio de los errores se mantuvo más contenido, con valores mayormente entre 500 y 750 MW, aunque se evidencian nuevos aumentos hacia el final del periodo, destacando el mes de enero de 2024 con más de 1000 MW. Estos resultados sugieren que, si bien el sistema eléctrico mostró signos de estabilización luego del fin de la pandemia, persisten eventos recientes que generan desviaciones significativas en la demanda, posiblemente vinculados a factores externos o estructurales del sistema.

Figura 4.20

Distribución de anomalías por mes y promedio de amplitud por mes



(a)



(b)

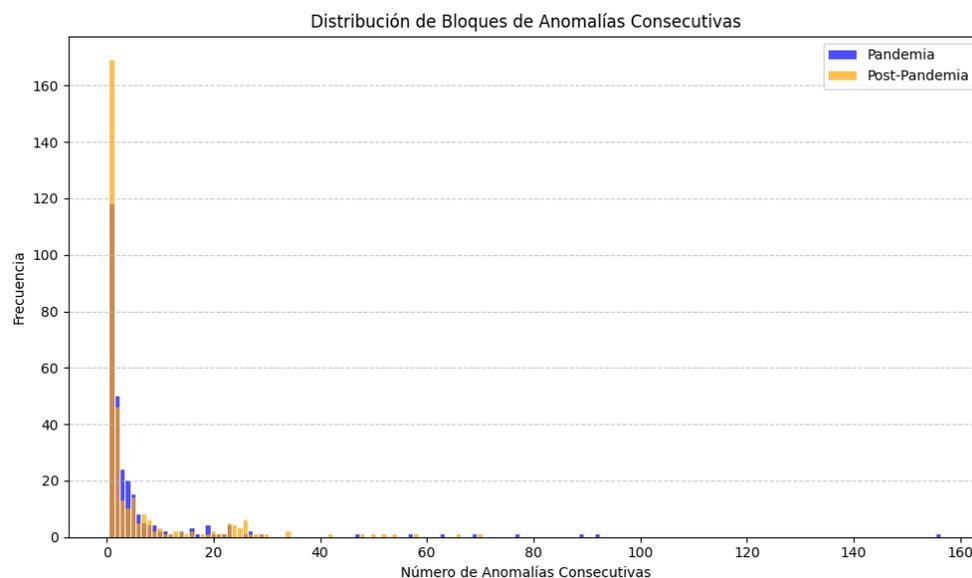
Fuente: Elaboración propia.

En la Figura 4.21 muestra la distribución de bloques de anomalías consecutivas durante la pandemia (en azul) y post-pandemia (en amarillo). Este análisis destaca cómo se agrupan las anomalías en ambos periodos y sugiere diferencias en la estabilidad del sistema. En ambos periodos, los bloques de anomalías más comunes son aquellos con pocas anomalías consecutivas (bloques de 1 o 2 anomalías). Sin embargo, post-pandemia se observa un número significativamente mayor de bloques cortos en comparación con la pandemia, lo que podría indicar un aumento en la frecuencia de eventos pequeños o transitorios. Durante la pandemia, se identifican algunos bloques largos (más de 40 anomalías consecutivas), aunque son raros. Esto refleja eventos sos-

tenidos de inestabilidad, posiblemente causados por cambios abruptos en la demanda. Post-pandemia, los bloques largos son mucho menos frecuentes, lo que sugiere que la demanda eléctrica se ha estabilizado y es menos propensa a eventos prolongados fuera de lo normal. La menor cantidad de bloques largos en el periodo post-pandemia podría indicar una recuperación hacia patrones más normales de demanda eléctrica. Aunque hay más bloques cortos, esto puede ser resultado de fluctuaciones menores y menos significativas. La pandemia presenta una distribución más dispersa con bloques de mayor longitud, mientras que post-pandemia la mayoría de las anomalías se concentran en bloques cortos, lo que refuerza la idea de una estabilización parcial. En conclusión en el caso de post-pandemia, las anomalías consecutivas tienden a ser más cortas y menos frecuentes en bloques largos, lo que sugiere una mayor estabilidad en la demanda eléctrica. Sin embargo, el aumento de bloques cortos podría indicar la presencia de pequeñas variaciones transitorias que aún deben ser monitoreadas. Este cambio refleja una normalización general, pero también sugiere la necesidad de analizar los factores que generan estas anomalías breves.

Figura 4.21

Comparación de anomalías consecutivas etapas durante y post-pandemia



Fuente: Elaboración propia.

Análisis Figura 4.22a frecuencia de Anomalías por Hora, muestra la cantidad

de anomalías detectadas en cada hora del día, separando los periodos de pandemia (azul) y post-pandemia (amarillo). Donde se puede observar:

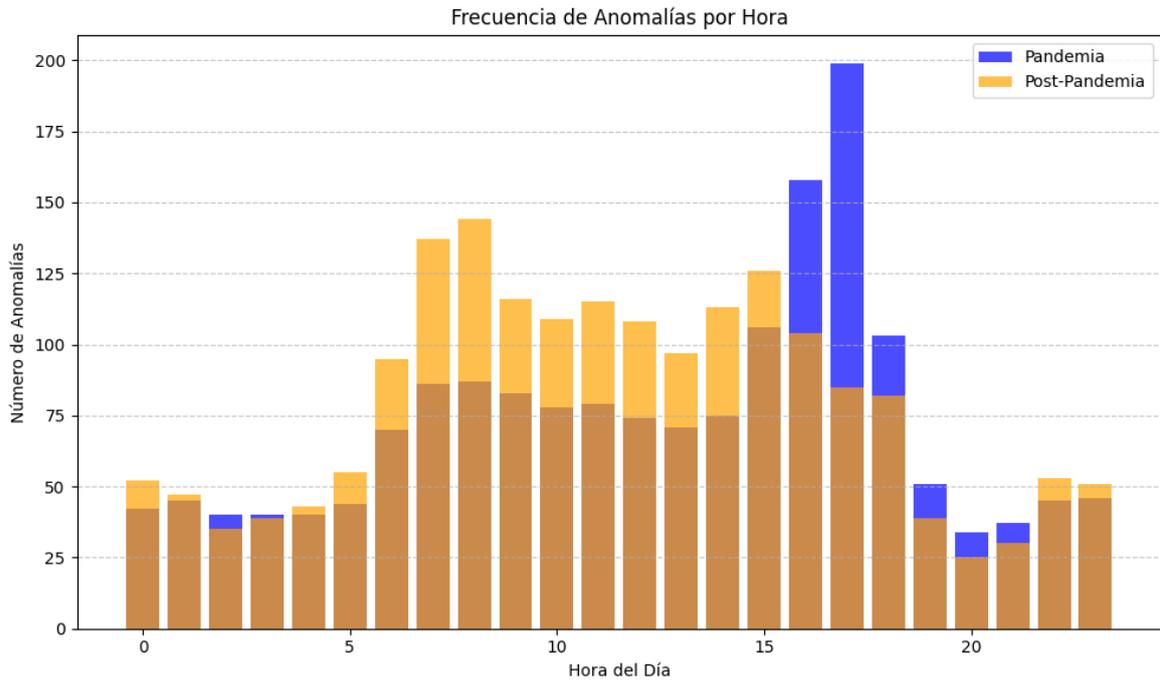
- Patrón durante la pandemia: Las anomalías se concentran significativamente entre las 15:00 y 16:00 horas, posiblemente reflejando un comportamiento atípico en la demanda eléctrica durante las horas pico. Fuera de este rango horario, las anomalías son menos frecuentes.
- Patrón post-pandemia: Hay un aumento generalizado de anomalías en casi todas las horas del día, especialmente en horas matutinas (05:00 a 10:00) y durante las horas de la tarde (15:00 a 18:00). La dispersión de anomalías a lo largo del día sugiere un cambio en los patrones de consumo post-pandemia, con menos concentraciones en horarios específicos.
- Comparación clave: La mayor frecuencia de anomalías en la tarde durante la pandemia contrasta con la distribución más uniforme post-pandemia, lo que podría indicar una estabilización parcial con eventos más pequeños pero distribuidos.

En la Figura 4.22b: Horarios de Inicio de Anomalías Consecutivas este gráfico destaca los horarios en los que comienzan bloques consecutivos de anomalías.

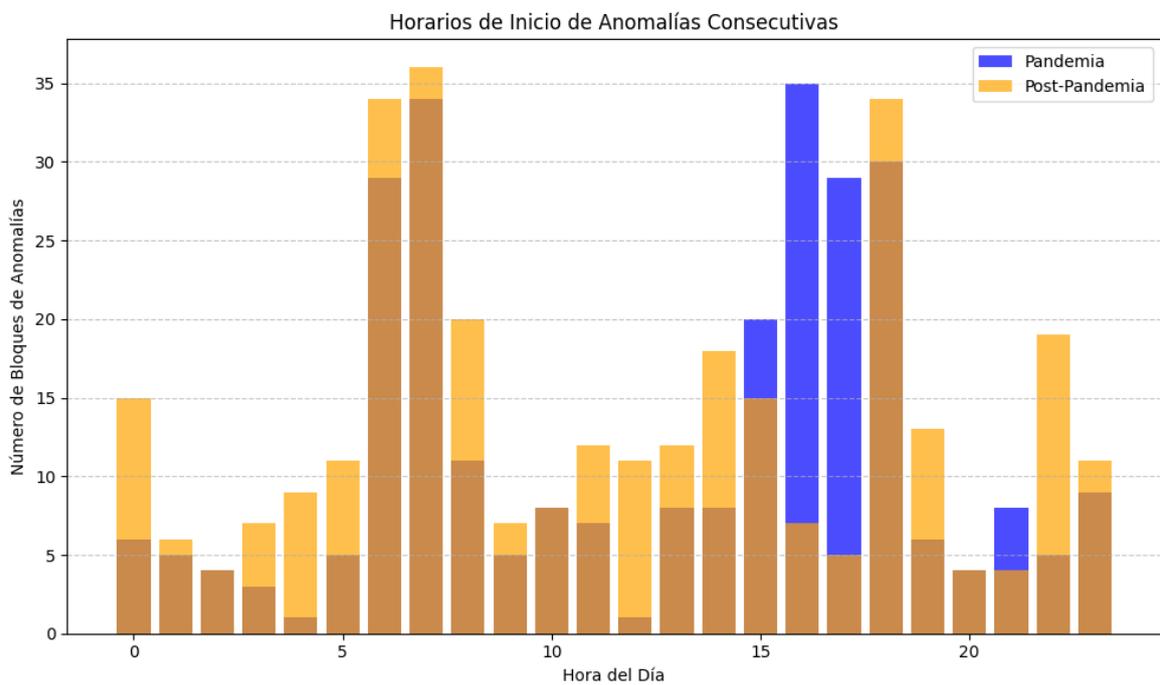
- Pandemia: Los bloques consecutivos inician predominantemente entre las 15:00 y 16:00 horas, reforzando la idea de que las anomalías más persistentes se asocian a este rango horario.
- Post-pandemia: Los horarios de inicio de bloques consecutivos son más variados, pero aún se observa un ligero pico entre las 04:00 y 05:00, además de un comportamiento notable entre las 15:00 y 18:00 horas.
- Diferencias clave: En la pandemia, los bloques consecutivos están fuertemente vinculados a un único rango horario (tarde). Post-pandemia, aunque el rango de la tarde sigue siendo relevante, aparecen más bloques en horarios dispersos, lo que refleja un cambio en los patrones de persistencia de anomalías.

Figura 4.22

Anomalías por hora puntual y consecutivas en los periodos durante y post-pandemia



(a)



(b)

Fuente: Elaboración propia.

Post-pandemia, las anomalías no solo aumentan en frecuencia a lo largo del

día, sino que también muestran una mayor dispersión en los horarios en los que inician bloques consecutivos. Esto sugiere que, aunque los patrones de consumo se han estabilizado en cierta medida, los eventos anómalos ocurren de manera más distribuida, reflejando posibles nuevos comportamientos en la demanda eléctrica. Este cambio podría estar relacionado con dinámicas laborales o hábitos de consumo modificados tras la pandemia.

4.7. Evaluación de resultados

Para evaluar los resultados relacionados con las anomalías detectadas en la serie temporal de demanda de energía eléctrica, se empleó el desempeño del modelo entrenado en la etapa pre-pandemia, aplicándolo a las etapas durante y post-pandemia. Como métricas de evaluación se utilizaron el RMSE (Raíz del error Cuadrático Medio), el MAE (Error Absoluto Medio) y el R^2 (Coeficiente de Determinación).

Tabla 4.4

Métricas con datos durante y post-pandemia.

Métrica	Durante	Post-pandemia
RMSE (normalizado)	0.0316	0.0296
RMSE (unidades originales)	102.6576	91.2199
RMSE %	1.72	1.36
MAE	73.2364	70.1181
R^2	0.9826	0.9762

Teniendo como referencia las estadísticas de la etapa pre-pandemia, donde la media es de 6,224 MW, la desviación estándar de 730 MW, el mínimo de 3,249 MW y el máximo de 8,304 MW, se interpreta a continuación la información presentada en la Tabla 4.4:

- RMSE (Raíz del Error Cuadrático Medio): Esta métrica mide la magnitud promedio de los errores de predicción. Cuando los datos están normalizados y el RMSE es menor a 0.05, el modelo se considera bueno. En este caso, utilizando

datos en su escala original, se obtienen valores de 102.66 MW y 91.23 MW para las etapas durante y post-pandemia, respectivamente. Estas cifras representan aproximadamente el 1.72 % y el 1.36 % de la media (6,224 MW), por lo que el error es muy pequeño en relación con el promedio de la demanda. Además, se observa que el RMSE disminuye en la etapa post-pandemia, lo que implica que el comportamiento del consumo se estabiliza con el tiempo. Estos valores indican que el modelo posee una alta precisión, con errores relativos bajos frente a la magnitud de la demanda energética.

- MAE (Error Absoluto Medio): El MAE mide el error promedio absoluto entre las predicciones y los valores reales. Durante la pandemia se obtiene un MAE de 73.23 MW, mientras que en la etapa post pandemia es de 70.11 MW. Estos valores equivalen a aproximadamente el 1.17 % de la media (6,224 MW), lo que sugiere un nivel de error aceptable. Asimismo, se aprecia una disminución del MAE en la etapa post pandemia, lo que indica que el comportamiento del consumo se ha estabilizado y el modelo mantiene una buena capacidad de predicción.
- R^2 (Coeficiente de Determinación): El coeficiente de determinación indica la proporción de la variabilidad de los datos explicada por el modelo. Dado que los valores de R^2 en ambas etapas (durante y post-pandemia) son cercanos a 1, puede interpretarse que el modelo explica la mayor parte de la variabilidad observada en la demanda. Esto respalda la conclusión de que el modelo es adecuado.

Dado que las métricas de error son bajas y el coeficiente de determinación es elevado, se concluye que el modelo presenta un buen desempeño en la predicción de la demanda de energía eléctrica tanto durante como después de la pandemia. La ligera mejora en las métricas post-pandemia sugiere que el modelo ha mantenido o incluso incrementado su precisión en condiciones cambiantes.

4.8. Análisis adicional

Se realizó un análisis con datos no considerados inicialmente en este estudio, abarcando el periodo desde el 14 de marzo hasta el 16 de diciembre de 2024. El objetivo fue verificar si el rendimiento del modelo mejoraba o empeoraba al incorporar nuevas observaciones. Los resultados se muestran en la Tabla 4.5:

Tabla 4.5

Métricas con datos de marzo a diciembre de 2024.

Métrica	Resultado
RMSE (normalizado)	91.2494
MAE	71.0539
R^2	0.9721

Estos valores confirman que el modelo sigue ofreciendo un buen desempeño, incluso al incorporar datos adicionales no contemplados en el análisis previo.

Con los datos adicionales, se comparó el rendimiento del modelo desarrollado en este estudio con las predicciones diarias y semanales proporcionadas por la plataforma del Comité de Operación Económica del Sistema (COES), que publica estimaciones de la demanda eléctrica. Los resultados se muestran en la Tabla 4.6 y en la Figura 4.23 se puede apreciar el comportamiento de las diferentes series temporales:

Tabla 4.6

Comparación de métricas de este estudio con datos predecidos del COES.

Métrica	Modelo	Prog. Diaria	Prog. Semanal
RMSE (unidades originales)	91.2494	127.0932	151.3013
MAE	71.0539	100.2981	117.1248
R^2	0.9721	0.9462	0.9237

Al comparar las métricas de evaluación, se observa que el modelo desarrollado en este trabajo presenta menor error y menor variabilidad que las proyecciones diarias y semanales del COES, lo que indica un mejor desempeño.

1 # Calcular el RMSE

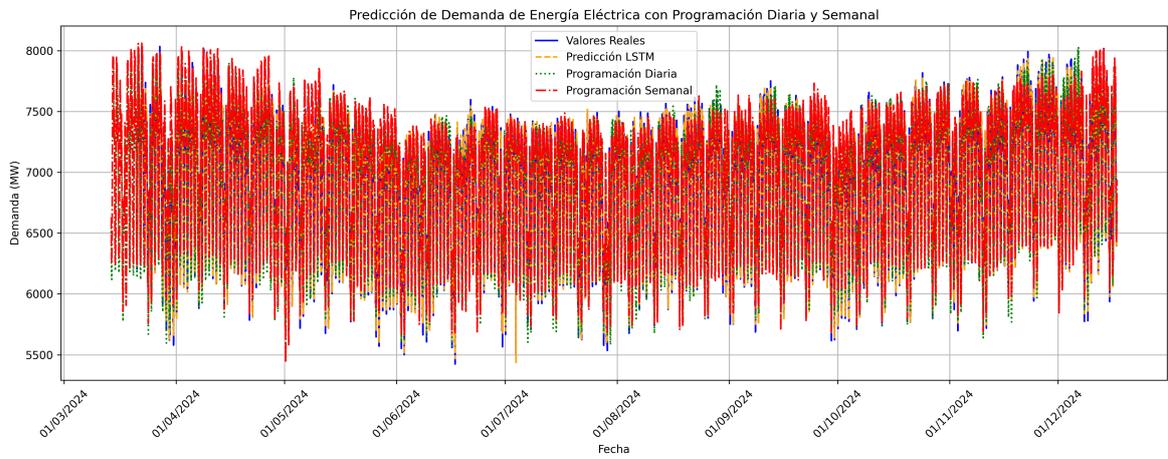
```

2 rmse_prog_diaria = np.sqrt(mean_squared_error(data_nueva['
      EJECUTADO'], data_nueva['PROG. DIARIA']))
3 print(rmse_prog_diaria)
4 rmse_prog_semanal = np.sqrt(mean_squared_error(data_nueva['
      EJECUTADO'], data_nueva['PROG. SEMANAL']))
5 print(rmse_prog_semanal)
6     verbose=2)
7 # Calcular el MAE
8 mae_prog_diaria = mean_absolute_error(data_nueva['EJECUTADO
      '], data_nueva['PROG. DIARIA'])
9 print(mae_prog_diaria)
10 mae_prog_semanal = mean_absolute_error(data_nueva['
      EJECUTADO'], data_nueva['PROG. SEMANAL'])
11 print(mae_prog_semanal)
12 # Calcular el coeficiente de determinacion
13 r2_prog_diaria = r2_score(data_nueva['EJECUTADO'],
      data_nueva['PROG. DIARIA'])
14 print(r2_prog_diaria)
15 r2_prog_semanal = r2_score(data_nueva['EJECUTADO'],
      data_nueva['PROG. SEMANAL'])
16 print(r2_prog_semanal)

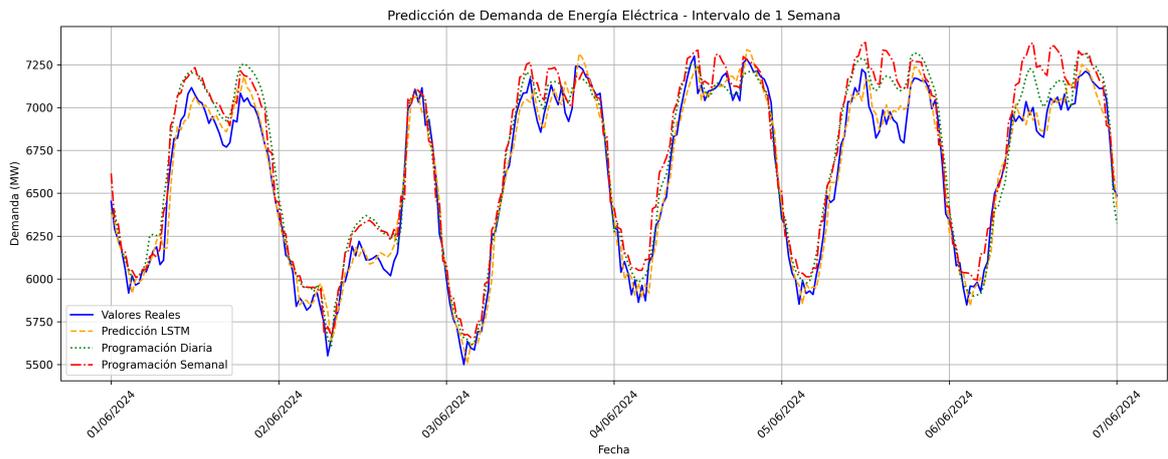
```

Figura 4.23

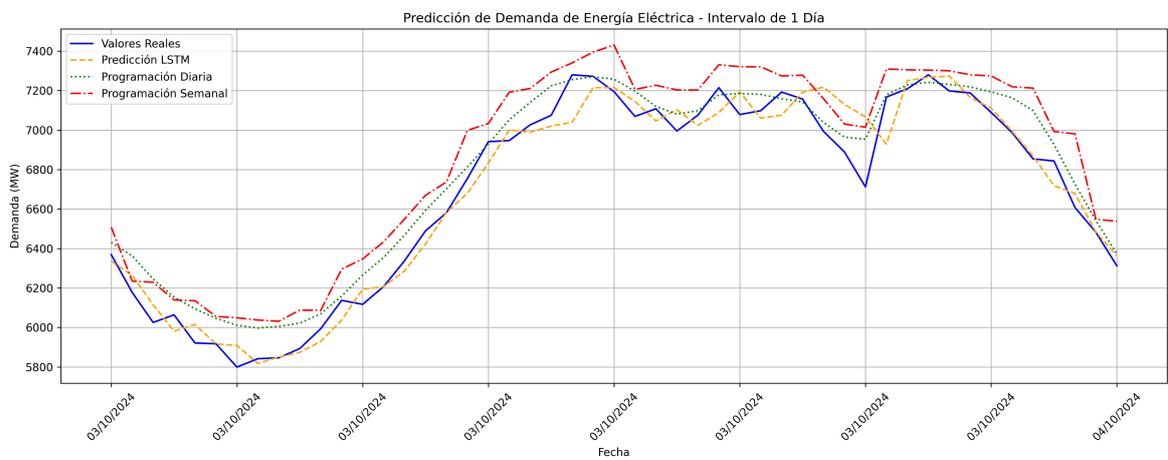
Comparación de valores reales, predichos por el modelo y valores programados por el COES



(a)



(b)



(c)

Fuente: Elaboración propia.

La Figura 4.23 muestra una comparación de la demanda de energía eléctrica real frente a tres enfoques predictivos: el modelo LSTM propuesto en esta investigación, y los valores proporcionados por el Comité de Operación Económica del Sistema (COES), específicamente las programaciones diaria y semanal. En la primera imagen, correspondiente a un periodo extendido (varios meses), es difícil apreciar las diferencias debido al alto nivel de superposición, por lo que se presentan otras dos vistas con intervalos más específicos para mayor claridad. En la segunda imagen, con un intervalo semanal, y especialmente en la tercera imagen, que muestra el comportamiento durante un día específico, se puede apreciar claramente que el modelo LSTM propuesto (línea amarilla discontinua) realiza predicciones más cercanas a los valores reales (línea azul continua), mientras que la programación diaria (línea verde punteada) y la programación semanal (línea roja discontinua) exhiben mayores desviaciones respecto a los valores observados. Esto demuestra que el modelo LSTM desarrollado es más preciso para predecir la dinámica de la demanda energética, especialmente al capturar cambios sutiles y patrones diarios que los modelos tradicionales del COES no logran representar con la misma exactitud.

Capítulo 5

Discusión de resultados

Los resultados obtenidos a lo largo de este estudio confirman que las redes neuronales LSTM constituyen una herramienta eficaz para modelar y predecir series temporales complejas como la demanda de energía eléctrica en el Perú. En este capítulo se discuten estos hallazgos, se comparan con investigaciones previas, y se plantean sus implicancias y limitaciones.

5.1. Rendimiento del modelo predictivo

El modelo LSTM entrenado mostró un buen desempeño predictivo, particularmente cuando se utilizaron ventanas de entrada que respetaban los ciclos horarios y semanales del consumo energético (24, 48 y 336 pasos de tiempo). Las métricas de evaluación, como la raíz del error cuadrático medio (RMSE), el error absoluto medio (MAE) y el coeficiente de determinación (R^2), evidenciaron que el modelo logró capturar adecuadamente la dinámica de la serie temporal.

Estos resultados coinciden con lo reportado por Fischer and Krauss (2018) y Kratzert et al. (2022), quienes señalan que las LSTM son altamente eficaces en contextos con estacionalidad, ruido y dependencias temporales prolongadas.

5.2. Identificación de anomalías

Una de las principales contribuciones del estudio fue la identificación de anomalías en la demanda energética durante y después de la pandemia de COVID-19. El uso del percentil 95 de los errores absolutos como umbral permitió detectar puntos con comportamiento atípico, mayormente concentrados en los años 2020 y 2021.

Este hallazgo es consistente con estudios como Zhao et al. (2021) y Fernando et al. (2020), que documentan cómo eventos disruptivos como la pandemia generan alteraciones estructurales en patrones de consumo. Aunque la demanda energética pareció volver a sus valores previos en el periodo post-pandemia, las anomalías persistentes podrían indicar cambios en el estilo de vida, como el teletrabajo o el comercio digital.

5.3. Implicancias para la planificación energética

Los resultados obtenidos pueden ser de gran utilidad para entidades como el COES, encargadas de planificar y gestionar el sistema eléctrico nacional. Un modelo predictivo robusto no solo permite anticipar la demanda, sino también detectar desviaciones inusuales en tiempo real, lo que es crucial para prevenir fallos en el suministro, reducir costos y mejorar la eficiencia del sistema.

5.4. Limitaciones del estudio

Aunque los resultados fueron positivos, existen algunas limitaciones. El modelo fue entrenado con datos históricos sin incorporar variables externas como temperatura, actividad económica o días feriados, que podrían influir en la demanda. Además, la detección de anomalías fue basada únicamente en el error del modelo, sin una validación

externa o supervisada.

5.5. Propuestas futuras

Para estudios futuros se sugiere:

- Incorporar variables exógenas para enriquecer el modelo.
- Explorar técnicas híbridas LSTM con atención o autoencoders.
- Validar las anomalías detectadas con fuentes externas (por ejemplo, medidas de confinamiento u horarios de toque de queda).

5.6. Comparación con estudios del estado del arte

Diversos estudios recientes han abordado la predicción de la demanda eléctrica y la detección de anomalías en series temporales utilizando redes neuronales LSTM o modelos híbridos. A continuación, se destacan algunos de los más relevantes por su similitud metodológica y valor comparativo con el presente trabajo:

- Dinh et al. (2024) compararon una red Bidireccional LSTM (BiLSTM) con un modelo híbrido CNN-LSTM para predecir la demanda energética en series temporales univariadas. Los resultados mostraron que el modelo CNN-LSTM superó significativamente al BiLSTM en todas las métricas (RMSE, MAE, MAPE y R^2), validando la utilidad de arquitecturas híbridas para mejorar la precisión predictiva. Este enfoque respalda el uso de LSTM y sugiere la posibilidad de ampliar el presente trabajo hacia modelos híbridos.
- Rahman et al. (2024) desarrollaron un sistema combinado de predicción y detección de anomalías mediante LSTM e Isolation Forest, aplicado al consumo

energético residencial. El modelo logró un RMSE de 0.0799 y una tasa perfecta de detección de anomalías (100%). Este enfoque confirma la viabilidad del análisis no supervisado adoptado en esta tesis y sugiere que integrar modelos complementarios podría mejorar aún más la identificación de patrones atípicos.

- Hanif et al. (2021) propusieron una arquitectura CNN-LSTM que mejoró en un 29% la raíz error cuadrático medio frente al modelo LSTM estándar. También se logró identificar de manera automática apagones y consumos inusuales. Estos resultados validan el enfoque basado en el error de predicción utilizado en esta tesis para detectar anomalías.
- Blázquez G. et al. (2021) realizaron una revisión exhaustiva sobre métodos de detección de outliers en series temporales, resaltando la falta de estandarización en la terminología y la necesidad de abordar la detección en contextos multivariantes. Esta revisión fundamenta teóricamente la necesidad de un análisis riguroso y específico en trabajos como el presente.
- Ergen and Kozat (2019) introdujeron técnicas no supervisadas y semisupervisadas de detección de anomalías con LSTM, mostrando superioridad respecto a métodos convencionales. Además, destacaron el uso de funciones de puntuación como SVDD y OC-SVM. Su flexibilidad para secuencias de longitud variable es relevante para este estudio, dado el volumen y granularidad de los datos analizados.
- Niu et al. (2020) combinaron LSTM con VAE-GAN para detectar anomalías en procesos industriales, mostrando mejoras en velocidad y precisión. Su enfoque fortalece el marco metodológico empleado en esta tesis y sugiere posibles ampliaciones con técnicas generativas.
- Bakhtawar M. (2019) analizó la detección de anomalías con LSTM en comparación con redes FFNN, destacando su capacidad para capturar patrones diarios y semanales en la demanda eléctrica. Aunque no siempre presentó superioridad estadísticamente significativa, orientó sobre la importancia de considerar la na-

turalidad del patrón al elegir el modelo.

- Malhotra et al. (2015) validaron el uso de redes LSTM apiladas para detectar patrones a largo plazo y anomalías mediante la distribución de errores. Este estudio refuerza el enfoque de esta investigación, que emplea umbrales sobre los errores absolutos para identificar valores atípicos.
- Pan et al. (2022) utilizaron una combinación de CNN, BiLSTM y mecanismos de atención para detectar anomalías en datos energéticos de alta dimensión. Sus resultados evidencian que la combinación de componentes puede mejorar el desempeño, lo cual representa una oportunidad futura para fortalecer el modelo de esta tesis.
- Finalmente, Cui et al. (2023) propusieron un marco de predicción ajustado al contexto de la pandemia COVID-19, integrando regresión lineal y LSTM multi-escalar sobre residuos. Su enfoque demostró mejoras superiores al 17% frente a modelos tradicionales, ofreciendo una base metodológica sólida para analizar el impacto de la pandemia, en línea con los objetivos de este estudio.

En conjunto, estos trabajos respaldan la elección metodológica de esta tesis, y ofrecen perspectivas valiosas para su validación, comparación y evolución futura. Asimismo, refuerzan el uso de LSTM y métodos de análisis de errores para la identificación de anomalías prolongadas en contextos de cambio disruptivo como el generado por la pandemia.

5.7. Comparación de resultados con el estado del arte

La Tabla 5.1 presenta una síntesis comparativa entre los resultados obtenidos en el presente estudio y los reportados en investigaciones recientes sobre predicción y detección de anomalías en series temporales de consumo energético.

Los trabajos de Dinh et al. (2024) y Hanif et al. (2021) utilizaron arquitecturas híbridas CNN-LSTM, logrando valores de error RMSE superiores a los obtenidos en esta investigación. Sin embargo, dichas investigaciones emplearon datos horarios o minutos de consumo, diferentes ventanas de entrada y contextos geográficos distintos, lo cual puede explicar la magnitud de los errores obtenidos. A pesar de ello, sus resultados respaldan la efectividad de las redes LSTM en tareas de predicción energética.

Por otro lado, el estudio de Rahman et al. (2024) propuso un enfoque híbrido basado en LSTM e Isolation Forest, logrando una excelente precisión en la detección de anomalías (F1-score = 1.0). Esto valida la estrategia empleada en la presente tesis, que también compara valores predichos y reales para identificar eventos atípicos, aunque desde un enfoque no supervisado basado en percentiles.

El presente trabajo, al emplear una red LSTM univariada sobre una serie de tiempo de seis años en el contexto peruano, logró un RMSE de 18.12 y un MAE de 13.27. Estos valores evidencian un buen ajuste del modelo en la predicción de la demanda energética. Asimismo, el método de detección de anomalías basado en el percentil 95 de los errores permitió identificar eventos fuera del comportamiento normal sin recurrir a modelos supervisados externos, lo que representa una solución más sencilla y aplicable en contextos con datos limitados etiquetados.

En conjunto, esta comparación demuestra que el enfoque adoptado en esta investigación es competitivo, y que su simplicidad metodológica no compromete la precisión. Además, proporciona una base sólida para futuras mejoras, como la incorporación de variables exógenas o el uso de modelos híbridos más complejos.

Tabla 5.1*Comparación de resultados con estudios del estado del arte*

Estudio	Modelo	RMSE	MAE	Detección de anomalías
Dinh et al. (2024)	CNN-LSTM	671.37	499.08	No aplica
Rahman et al. (2024)	LSTM Isolation Forest	0.0799	–	Precisión, Recall, F1 = 1.0
Hanif et al. (2021)	CNN-LSTM	2097.00	–	Basado en MAE y umbral
Estudio presente	LSTM univariado	18.12	13.27	Basado en percentil 95 del error

Nota. Elaboración propia a partir de los resultados reportados en los estudios citados.

Capítulo 6

Conclusiones y recomendaciones

6.1. Conclusiones

- Se desarrolló exitosamente un modelo basado en redes neuronales recurrentes tipo LSTM, entrenado con datos históricos pre-pandemia. El modelo propuesto logró captar adecuadamente la dinámica temporal y los patrones semanales característicos de la demanda eléctrica, demostrando así su capacidad para modelar series temporales complejas con comportamientos cíclicos.
- El modelo mostró un desempeño destacado, alcanzando una raíz del error cuadrático medio (RMSE) inferior al 2% de la demanda promedio y un error absoluto medio (MAE) reducido, así como un coeficiente de determinación (R^2) superior al 97%. Estos resultados evidencian la alta precisión y efectividad del modelo LSTM para predecir la demanda eléctrica en contextos normales y también bajo condiciones disruptivas.
- Sobre la identificación y análisis de anomalías: El modelo identificó claramente anomalías durante y después de la pandemia al comparar valores predichos con los observados. Durante la pandemia se detectaron 1,733 anomalías asociadas principalmente con restricciones sanitarias y económicas. En el periodo post-

pandemia, se identificaron 1,916 anomalías, lo que evidencia una persistencia significativa de eventos atípicos. Este aumento de anomalías puede atribuirse a la inestabilidad política y social del país, así como a patrones cambiantes en la recuperación económica.

6.2. Recomendaciones

- Se recomienda explorar la inclusión de variables adicionales como factores climáticos (temperatura, precipitación, humedad), indicadores económicos (variación del PBI, precios energéticos), eventos políticos y sociales específicos (paros, protestas, elecciones), así como variables relacionadas con movilidad y teletrabajo, a fin de evaluar si estas variables mejoran aún más la capacidad predictiva y explicativa del modelo.
- Considerando la diversidad geográfica y socioeconómica del Perú, se recomienda aplicar el modelo LSTM por regiones (costa, sierra, selva) o por ciudades principales, para determinar si existen patrones diferenciados de consumo energético. Esto permitiría desarrollar modelos más precisos y adaptados a contextos regionales específicos, fortaleciendo las estrategias locales de gestión energética.
- Es recomendable evaluar modelos híbridos que combinen LSTM con técnicas complementarias como Isolation Forest, autoencoders o métodos estadísticos avanzados, para aumentar la precisión en la detección y clasificación automática de anomalías. Esto facilitaría el monitoreo inteligente del sistema eléctrico nacional, mejorando la identificación temprana de eventos críticos.
- Dado que los patrones de demanda energética pueden variar continuamente, se recomienda desarrollar un sistema automatizado que permita la actualización periódica (incluso diaria o semanal) del modelo LSTM con nuevos datos, garantizando su adaptabilidad a cambios recientes en la demanda eléctrica y manteniendo una alta precisión predictiva a largo plazo.
- Se recomienda evaluar la integración práctica del modelo desarrollado con los sistemas existentes en entidades reguladoras como el COES, para realizar pruebas piloto en escenarios reales. Esto permitirá evaluar su eficacia práctica, así como mejorar su interfaz, capacidad de interpretación y utilidad en la toma de decisiones operativas y estratégicas.
- Ante el crecimiento proyectado de nuevas tecnologías, como vehículos eléctricos

y generación distribuida (paneles solares, sistemas de almacenamiento), es recomendable realizar estudios adicionales sobre cómo estos cambios pueden influir en los patrones futuros de consumo eléctrico. Esto ofrecerá insumos valiosos para la planificación estratégica del sector energético nacional.

- Se sugiere realizar estudios específicos para evaluar la capacidad del modelo propuesto para anticipar y manejar eventos extremos como crisis energéticas globales, desastres naturales o nuevas crisis sanitarias. Este análisis permitirá conocer la verdadera robustez del modelo, ayudando a definir protocolos claros de respuesta ante emergencias.

Bibliografía

- Ahmed, M., Mahmood, A. N., and Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31.
- Al-amri, R., Murugesan, R. K., Man, M., Abdulateef, A. F., Al-Sharafi, M. A., and Alkahtani, A. A. (2021). A review of machine learning and deep learning techniques for anomaly detection in iot data. *Applied Sciences*, 11(12):5320.
- Alonso, G. (2012). Análisis de la demanda de energía eléctrica necesaria para posibilitar la industrialización de productos agrícolas en el departamento de itapúa. *Universidad Nacional de Itapúa*.
- Bakhtawar M., S. (2019). Anomaly detection in electricity demand time series data. School of Electrical Engineering and Computer Science.
- Blázquez G., A., Conde, A., Mori, U., and Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR)*, 54(3):1–33.
- Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2016). *Time Series Analysis - Forecasting and Control*. Wiley, USA.
- Brownlee, J. (2018). *Deep Learning for Time Series Forecasting*. Machine Learning Mastery.
- Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):1–58.
- Chatfield, C. and Xing, H. (2019). *The Analysis of time series*. CRC Press, USA.

- Cissoko, M. B. H. (2024). *Adaptive time-aware LSTM for predicting and interpreting ICU outcomes*. PhD thesis, Université Paris-Saclay.
- COES (2024). Demanda energía eléctrica Perú. URL: <https://www.coes.org.pe/Portal/home/>. Recuperado el 4 de septiembre de 2024.
- Cui, Z., Wu, J., Lian, W., and Wang, Y.-G. (2023). A novel deep learning framework with a covid-19 adjustment for electricity demand forecasting. *Energy Reports*, 9:1887–1895.
- Dinh, P. N., Nguyen, Q. D., and Doan, M. D. (2024). Electrical energy demand forecasting using time series in lstm and cnn-lstm. *Journal of Industrial and Systems Engineering Management (JISEM)*, 5(1):73–83. Accedido en julio de 2025.
- EC, R. (2020). Coronavirus en Perú: “vamos a mantener la calma y confiar en el sistema de salud”, dice Martín Vizcarra. *El Comercio*. URL: <https://elcomercio.pe/peru/coronavirus-en-peru-martin-vizcarra-confirma-primer-caso-del-covid-protect\penalty-\@M-19-en-el-pais-nndc-noticia/?ref=ecr>. Recuperado el 1 de septiembre de 2024.
- EP, R. (2020). Gobierno declara estado de emergencia nacional y aislamiento social obligatorio por 15 días. *El Peruano*. URL: <https://elperuano.pe/noticia/92075-gobierno-declara-estado-de-emergencia-nacional-y-aislamiento-protect\penalty-\@M-social-obligatorio-por-15-dias>. Recuperado el 1 de septiembre de 2024.
- Ergen, T. and Kozat, S. S. (2019). Unsupervised anomaly detection with lstm neural networks. *IEEE transactions on neural networks and learning systems*, 31(8):3127–3141.
- Fernando, T., Gammulle, H., Denman, S., Sridharan, S., and Fookes, C. (2020). Deep learning for medical anomaly detection – a survey. *arXiv preprint arXiv:2012.02364*.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications, 4th edition.

- Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669.
- GeeksforGeeks (2025). Introducción a long short-term memory (lstm). <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>. Accedido el: 22-04-2025.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2nd edition.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.
- Hanif, M., Ullah, I., and Rehman, A. U. (2021). Power consumption predictive analytics and automatic anomaly detection using cnn-lstm. In *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–6. IEEE. Accedido en julio de 2025.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition.
- Hernández Sampieri, R., Fernández Collado, C., and Baptista Lucio, P. (2014). *Metodología de la Investigación*. Mc Graw Hill, Mexico.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Kratzert, F. et al. (2022). Using a long short-term memory (lstm) neural network to boost streamflow forecasting. *Hydrology and Earth System Sciences*, 26(20):5449–5468.

- Limouni, T., Yaagoubi, R., Bouziane, K., Guissi, K., and Baali, E. H. (2022). Univariate and multivariate lstm models for one-step and multistep pv power forecasting. *International Journal of Renewable Energy Development*, 11(3):815–828.
- Malhotra, P., Vig, L., Shroff, G., Agarwal, P., et al. (2015). Long short term memory networks for anomaly detection in time series. In *Esann*, volume 2015, page 89.
- MINSA (2022). Gobierno oficializa fin del estado de emergencia por la covid-19. *gob.pe*. URL: <https://www.gob.pe/institucion/minsa/noticias/664735-gobierno-oficializa-fin-del-estado-de-emergencia-por-la-covid-19>. Recuperado el 1 de septiembre de 2024.
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc.
- Montgomery, D. C. and Runger, G. C. (2014). *Applied Statistics and Probability for Engineers*. Wiley, 6th edition.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.
- Niu, Z., Yu, K., and Wu, X. (2020). Lstm-based vae-gan for time-series anomaly detection. *Sensors*, 20(13):3738.
- OMS (2020). Coronavirus disease (covid-19). *World Health Organization*. URL: https://www.who.int/health-topics/coronavirus#tab=tab_1. Recuperado el 1 de septiembre de 2024.
- Osinergmin (2017). La industria de la electricidad en el Perú, 25 años de aportes al crecimiento económico del país. *Osinergmin*, 1(1):30–31. URL: https://www.osinergmin.gob.pe/seccion/centro_documental/Institucional/Estudios_Economicos/Libros/Osinergmin-Industria-Electricidad-Peru-25anos.pdf. Recuperado el 8 de septiembre de 2024.

- Pan, H., Yin, Z., and Jiang, X. (2022). High-dimensional energy consumption anomaly detection: A deep learning-based method for detecting anomalies. *Energies*, 15(17).
- Pang, G., Shen, C., Cao, L., and van den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, 54(2):1–38.
- Rahman, M. S., Amin, M. B. B., and Alvee, M. S. R. (2024). Time series-based analysis of energy consumption: Forecasting and anomaly detection using lstm and isolation forest. *International Journal of Engineering Research and Technology*, 13(5):631–639. Accedido en julio de 2025.
- Russell, S. J. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson, Upper Saddle River, NJ, 3rd edition.
- Teng, X., Lin, Y., and Wen, X. (2017). *Anomaly detection in dynamic networks using multi-view time-series hypersphere learning*,. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Wu, H.-S. (2016). A survey of research on anomaly detection for time series. *2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. URL: keywords=Anomaly detection;Time series analysis;Data models;Clustering algorithms;Algorithm design and analysis;Predictive models;Time series;anomaly detection;big data;data mining;multivariate time series, doi=10.1109/ICCWAMTIP.2016.8079887.
- Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2007). Long short-term memory (lstm). [recuperado el 14-04-2025].
- Zhao, Y. et al. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, 287:116601.

Anexos

Anexo 1: Librerías usadas

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from sklearn.preprocessing import MinMaxScaler
5 from tensorflow.keras.models import Sequential
6 from tensorflow.keras.layers import LSTM, Dense, Dropout
7 from tensorflow.keras.optimizers import RMSprop
8 import tensorflow as tf
```

Anexo 2: Carga, conversión y división de datos

```
1 # Configuración inicial
2 # Para llegar casi siempre al mismo resultado, definimos la
   semilla
3 tf.random.set_seed(123)
4 # Garantizar que la GPU tensorflow haga todo lo posible
   para que se entrene
5 # y se tenga resultados similares
6 tf.config.experimental.enable_op_determinism()
7
8 # 1. Cargar los datos y preprocesar
9 # Cargar archivo CSV
10 data = pd.read_csv('Datos_demanda.csv', sep=';')
11
```

```

12 # Seleccionar columnas relevantes y convertir FECHA a
    datetime
13 data['FECHA'] = pd.to_datetime(data['FECHA'], format='%d/%m
    /%Y %H:%M')
14 data.set_index('FECHA', inplace=True)
15 data = data[['EJECUTADO']]
16
17 # Dividir en fases temporales
18 pre_pandemia = data[data.index < '2020-03-14']
19 durante_pandemia = data[(data.index >= '2020-03-14') & (
    data.index <= '2022-03-13')]
20 post_pandemia = data[data.index > '2022-03-13']

```

Anexo 3: Análisis y gráficos iniciales

```

1 # Descomposicion de la serie pre-pandemia
2 result = seasonal_decompose(pre_pandemia['EJECUTADO'],
    model='additive', period=365)
3 result.plot()
4 plt.title('Descomposicion de la Demanda Pre-Pandemia')
5 plt.show()
6
7 # Estadísticas generales pre-pandemia
8 print("Estadísticas de la demanda pre-pandemia:")

```

```

1 import matplotlib.pyplot as plt
2
3 # Graficar la serie temporal completa con colores para cada
    fase

```

```

4 plt.figure(figsize=(10, 6), dpi=120)
5
6 # Fase pre-pandemia
7 plt.plot(pre_pandemia.index, pre_pandemia['EJECUTADO'],
8         label='Pre-pandemia', color='blue')
9
10 # Fase durante la pandemia
11 plt.plot(durante_pandemia.index, durante_pandemia['
12         EJECUTADO'], label='Durante pandemia', color='orange')
13
14 # Fase post-pandemia
15 plt.plot(post_pandemia.index, post_pandemia['EJECUTADO'],
16         label='Post-pandemia', color='green')
17
18 # Personalizar el grafico
19 plt.title('Demanda de Energia Electrica en Peru')
20 plt.xlabel('Fecha')
21 plt.ylabel('Demanda (MW)')
22 plt.legend()
23 plt.grid(True)
24
25 # Mostrar el grafico
26 plt.tight_layout()
27 plt.show()

```

```

1 # Filtrar los datos para graficar
2 data_dia = data.loc['2018-01-01':'2020-01-01']
3
4 Graficar solo ese dia

```

```

5 plt.figure(figsize=(10, 4), dpi=200)
6 plt.plot(data_dia.index, data_dia['EJECUTADO'], label='
    Demanda de Energia', color='blue')
7 plt.xlabel('Tiempo')
8 plt.ylabel('Demanda (MW)')
9 plt.grid(True)
10 plt.tight_layout()
11 plt.show()

```

Anexo 4: Normalización datos

```

1 # grafico de normalizacion
2 # Seleccionamos una parte para graficar
3 subset = data.loc['2021-01-01':'2021-02-28']
4 # Normalizar la serie temporal
5 scaler = MinMaxScaler(feature_range=(0, 1))
6 subset['EJECUTADO_ESCALADO'] = scaler.fit_transform(subset
    [['EJECUTADO']])
7 # Crear los sub graficos
8 fig, axs = plt.subplots(2, 1, figsize=(12, 8), dpi=120,
    sharex=True)
9 # Serie original
10 axs[0].plot(subset.index, subset['EJECUTADO'], label='Serie
    Original', color='blue')
11 axs[0].set_title('Demanda de Energia Electrica (Original)')
12 axs[0].set_ylabel('Demanda (MW)')
13 axs[0].grid(True)
14 axs[0].legend()
15 # Serie normalizada

```

```

16  axs[1].plot(subset.index, subset['EJECUTADO_ESCALADO'],
    label='Serie Normalizada (0 a 1)', color='green')
17  axs[1].set_title('Demanda de Energia Electrica (Normalizada
    )')
18  axs[1].set_ylabel('Demanda Escalada')
19  axs[1].set_xlabel('Fecha')
20  axs[1].grid(True)
21  axs[1].legend()
22
23  # Ajustar espaciado y mostrar
24  plt.tight_layout()
25  plt.show()

```

Anexo 5: Configuración de hiperparámetros y entrenamiento del modelo

```

1  # Preparar datos para entrenamiento y prediccionn
2  INPUT_LENGTH = 336 # 1 semana de datos
3  OUTPUT_LENGTH = 1 # la siguiente media hora
4
5  # Escalar datos pre-pandemia
6  pre_pandemia_scaled, scaler = escalar_dataset(pre_pandemia.
    values)
7
8  # Crear datasets supervisados para entrenamiento
9  x_tr, y_tr = crear_dataset_supervisado(pre_pandemia_scaled,
    INPUT_LENGTH, OUTPUT_LENGTH)
10 # 4. Crear y entrenar el modelo

```

```

11 N_UNITS = 256
12 INPUT_SHAPE = (x_tr.shape[1], x_tr.shape[2])
13
14 modelo = Sequential()
15 modelo.add(LSTM(N_UNITS, input_shape=INPUT_SHAPE))
16 #modelo.add(Dropout(0.1))
17 modelo.add(Dense(OUTPUT_LENGTH, activation='linear'))
18
19 def root_mean_squared_error(y_true, y_pred):
20     return tf.math.sqrt(tf.math.reduce_mean(tf.square(y_pred -
21         y_true)))
22
23 modelo.compile(
24     optimizer=RMSprop(learning_rate=5e-5),
25     loss=root_mean_squared_error)
26
27 modelo.summary()
28 EPOCHS = 50
29 BATCH_SIZE = 64
30
31 historia = modelo.fit(
32     x_tr,
33     y_tr,
34     epochs=EPOCHS,
35     batch_size=BATCH_SIZE,
36     validation_data = (x_tr, y_tr),
37     verbose=2)
38
39 #guardar modelo
40 modelo.save('modelo_demanda.h5')

```

Anexo 6: Métricas de evaluación del modelo

```
1 from sklearn.metrics import mean_squared_error,
    mean_absolute_error, r2_score
2 import numpy as np
3
4 # Predicciones durante la pandemia
5 y_pred_durante_pandemia_scaled = modelo.predict(
    x_durante_pandemia)
6 # Remodelar a 2D antes de invertir la escala
7 y_pred_durante_pandemia = scaler.inverse_transform(
    y_pred_durante_pandemia_scaled.reshape(-1, 1))
8 y_real_durante_pandemia = scaler.inverse_transform(
    y_durante_pandemia.reshape(-1, 1))
9
10 # Predicciones post pandemia
11 y_pred_post_pandemia_scaled = modelo.predict(
    x_post_pandemia)
12 # Remodelar a 2D antes de invertir la escala
13 y_pred_post_pandemia = scaler.inverse_transform(
    y_pred_post_pandemia_scaled.reshape(-1, 1))
14 y_real_post_pandemia = scaler.inverse_transform(
    y_post_pandemia.reshape(-1, 1))
15
16 # Durante la pandemia
17 mae_durante_pandemia = mean_absolute_error(
    y_real_durante_pandemia, y_pred_durante_pandemia)
18 r2_durante_pandemia = r2_score(y_real_durante_pandemia,
    y_pred_durante_pandemia)
```

```

19 rmse_durante_pandemia = np.sqrt(mean_squared_error(
    y_real_durante_pandemia, y_pred_durante_pandemia))
20 # Calcular RMSE% durante la pandemia
21 mean_real_durante_pandemia = np.mean(
    y_real_durante_pandemia)
22 rmse_percent_durante_pandemia = (rmse_durante_pandemia /
    mean_real_durante_pandemia) * 100
23
24 # Post pandemia
25 mae_post_pandemia = mean_absolute_error(
    y_real_post_pandemia, y_pred_post_pandemia)
26 r2_post_pandemia = r2_score(y_real_post_pandemia,
    y_pred_post_pandemia)
27 rmse_post_pandemia = np.sqrt(mean_squared_error(
    y_real_post_pandemia, y_pred_post_pandemia))
28 # Calcular RMSE% post pandemia
29 mean_real_post_pandemia = np.mean(y_real_post_pandemia)
30 rmse_percent_post_pandemia = (rmse_post_pandemia /
    mean_real_post_pandemia) * 100
31
32 # Resultados
33 print(f'RMSE durante pandemia: {rmse_durante_pandemia:.4f}'
    )
34 print(f'RMSE post pandemia: {rmse_post_pandemia:.4f}')
35 print(f'RMSE% durante pandemia: {
    rmse_percent_durante_pandemia:.2f}%')
36 print(f'RMSE% post pandemia: {rmse_percent_post_pandemia:.2
    f}%')
37 print(f'MAE durante pandemia: {mae_durante_pandemia:.4f}')

```

```

38 print(f'MAE post pandemia: {mae_post_pandemia:.4f}')
39 print(f'R_2 durante pandemia: {r2_durante_pandemia:.4f}')
40 print(f'R_2 post pandemia: {r2_post_pandemia:.4f}')

```

Anexo 7: Visualización de anomalías

```

1 plt.figure(figsize=(15, 5))
2 plt.plot(errores_pandemia, label='Errores durante pandemia'
3         )
4 plt.axhline(y=umbral, color='r', linestyle='--', label='
5         Umbral de anomalías')
6 plt.axhline(y=-umbral, color='r', linestyle='--')
7 plt.legend()
8 plt.title('Errores durante pandemia')
9 plt.show()
10
11 plt.figure(figsize=(15, 5))
12 plt.plot(errores_post_pandemia, label='Errores post
13         pandemia')
14 plt.axhline(y=umbral, color='r', linestyle='--', label='
15         Umbral de anomalías')
16 plt.axhline(y=-umbral, color='r', linestyle='--')
17 plt.legend()
18 plt.title('Errores post pandemia')
19 plt.show()

```

```

1 def predecir_visualizar(data, modelo, scaler, input_length,
2     output_length):
3     # Escalar los datos

```

```

3 data_scaled = scaler.transform(data.values)
4 # Crear el dataset supervisado
5 x, y = crear_dataset_supervisado(data_scaled, input_length,
    output_length)
6 # Predecir
7 y_pred = predecir(x, modelo, scaler)
8 # Crear un DataFrame con las predicciones
9 data_pred = data.copy()
10 data_pred['EJECUTADO_PRED'] = np.nan
11 data_pred.iloc[input_length:input_length+len(y_pred), 1] =
    y_pred
12 # Graficar
13 plt.figure(figsize=(15, 5))
14 plt.plot(data.index, data['EJECUTADO'], label='Demanda de
    Energia')
15 plt.plot(data_pred.index, data_pred['EJECUTADO_PRED'],
    label='Prediccion', linestyle='--')
16 plt.title('Demanda de Energia Electrica y Prediccion')
17 plt.xlabel('Fecha')
18 plt.ylabel('Demanda (MW)')
19 plt.legend()
20 plt.grid(True)
21 plt.tight_layout()
22 plt.show()
23
24 # Predecir y visualizar la demanda durante la pandemia
25 predecir_visualizar(durante_pandemia, modelo, scaler,
    INPUT_LENGTH, OUTPUT_LENGTH)
26 # Predecir y visualizar la demanda post-pandemia

```

```
27 predecir_visualizar(post_pandemia, modelo, scaler,  
    INPUT_LENGTH, OUTPUT_LENGTH)
```

Anexo 8: Análisis de anomalías

```
1 import matplotlib.pyplot as plt  
2  
3 errores_pandemia = np.random.normal(0, 1, 1000) # Replace  
    with actual error data for pandemic  
4 errores_post_pandemia = np.random.normal(0.1, 1, 1000) #  
    Replace with actual error data for post-pandemic  
5 umbral = 2 # Replace with your calculated threshold  
6  
7 plt.figure(figsize=(12, 6))  
8 plt.hist(errores_pandemia, bins=50, alpha=0.6, color='blue',  
    , label='Errores durante pandemia')  
9 plt.hist(errores_post_pandemia, bins=50, alpha=0.6, color='  
    green', label='Errores post pandemia')  
10 plt.axvline(x=umbral, color='red', linestyle='--', label='  
    Umbral de anomalias (+)')  
11 plt.axvline(x=-umbral, color='red', linestyle='--', label='  
    Umbral de anomalias (-)')  
12 plt.title('Distribucion de errores: durante la pandemia vs.  
    post pandemia')  
13 plt.xlabel('Error')  
14 plt.ylabel('Frecuencia')  
15 plt.legend()  
16 plt.grid(alpha=0.3)  
17 plt.show()
```

Anexo 9: Usar el modelo guardado y prueba con nuevos datos

```
1 import pandas as pd
2 import numpy as np
3 from sklearn.preprocessing import MinMaxScaler
4 import matplotlib.pyplot as plt
5 from tensorflow.keras.models import load_model
6 from sklearn.metrics import mean_squared_error,
    mean_absolute_error, r2_score
7 from sklearn.impute import SimpleImputer
8 import matplotlib.dates as mdates
9
10 # Parametros del modelo
11 INPUT_LENGTH = 336 # Ventana de tiempo (1 semana de datos)
12 OUTPUT_LENGTH = 1 # Predecir la siguiente media hora
13 # Funcion para preparar datos supervisados
14 def crear_dataset_supervisado(data, input_length,
    output_length):
15 X, Y = [], []
16 for i in range(len(data) - input_length - output_length +
    1):
17 X.append(data[i: i + input_length])
18 Y.append(data[i + input_length: i + input_length +
    output_length])
19 return np.array(X), np.array(Y)
20
21 # Cargar los datos desde el archivo CSV
```

```

22 file_path = "DemandaCOES_2024_1.csv"
23 data_nueva = pd.read_csv(file_path, sep=';')
24
25 # Convertir la columna de fechas al tipo datetime
26 data_nueva['FECHA'] = pd.to_datetime(data_nueva['FECHA'])
27
28 # Verificar si hay valores NaN en los datos y manejarlos
29 if data_nueva.isnull().values.any():
30     print("Datos contienen NaN. Procediendo a imputacion...")
31     # Imputar NaN con la media de la columna 'EJECUTADO'
32     imputer = SimpleImputer(strategy='mean')
33     data_nueva['EJECUTADO'] = imputer.fit_transform(data_nueva
34         [['EJECUTADO']])
35 else:
36     print("No se encontraron valores NaN en los datos.")
37
38 # Seleccionar la columna relevante y escalar los datos
39 scaler = MinMaxScaler(feature_range=(0, 1))
40 data_scaled = scaler.fit_transform(data_nueva[['EJECUTADO',
41     ]].values)
42
43 # Preparar los datos en formato supervisado
44 X_nuevo, Y_nuevo = crear_dataset_supervisado(data_scaled,
45     INPUT_LENGTH, OUTPUT_LENGTH)
46
47 # Ajustar las dimensiones de X_nuevo para que sean
48     compatibles con el modelo LSTM
49
50 X_nuevo = X_nuevo.reshape((X_nuevo.shape[0], X_nuevo.shape
51     [1], 1))

```

```

46
47 # Definir la funcion personalizada si fue utilizada durante
    el entrenamiento
48 import tensorflow.keras.backend as K
49 def root_mean_squared_error(y_true, y_pred):
50 return K.sqrt(K.mean(K.square(y_pred - y_true)))
51
52 # Cargar el modelo entrenado
53 modelo = load_model("modelo_demanda.h5", custom_objects={'
    root_mean_squared_error': root_mean_squared_error})
54
55 print("Modelo cargado correctamente.")
56
57 # Realizar predicciones
58 Y_pred_scaled = modelo.predict(X_nuevo)
59
60 # Ajustar las dimensiones de Y_pred_scaled y Y_nuevo para
    aplicar inverse_transform
61 Y_pred_scaled = Y_pred_scaled.reshape(-1, 1)
62 Y_nuevo = Y_nuevo.reshape(-1, 1)
63
64 # Invertir la escala de las predicciones y los valores
    reales
65 Y_pred = scaler.inverse_transform(Y_pred_scaled)
66 Y_real = scaler.inverse_transform(Y_nuevo)
67
68 # Calcular el RMSE
69 rmse = np.sqrt(mean_squared_error(Y_real, Y_pred))
70 print(f"Error Cuadratico Medio (RMSE): {rmse:.4f}")

```

```

71
72 # Calcular el MAE
73 mae = mean_absolute_error(Y_real, Y_pred)
74 print(f"Error Absoluto Medio (MAE): {mae:.4f}")
75
76 # Calcular el Coeficiente de Determinacion R_2
77 r2 = r2_score(Y_real, Y_pred)
78 print(f"Coeficiente de Determinacion (R_2): {r2:.4f}")
79
80 # Generar las fechas correspondientes para las predicciones
81 fechas_predicciones = data_nueva['FECHA'][INPUT_LENGTH +
      OUTPUT_LENGTH - 1:]
82
83 # Graficar las predicciones y los valores reales con fechas
      en el eje X
84 plt.figure(figsize=(15, 6))
85 plt.plot(fechas_predicciones, Y_real, label="Valores reales
      ", color="blue")
86 plt.plot(fechas_predicciones, Y_pred, label="Prediccion
      LSTM", color="orange", linestyle="--")
87 plt.title("Prediccion de Demanda de Energia Electrica")
88 plt.xlabel("Fecha")
89 plt.ylabel("Demanda (MW)")
90 plt.legend()
91 plt.grid()
92 plt.gca().xaxis.set_major_locator(mdates.AutoDateLocator())
93 plt.gca().xaxis.set_major_formatter(mdates.
      ConciseDateFormatter(mdates.AutoDateLocator()))
94 plt.xticks(rotation=45)

```

```

95 plt.tight_layout()
96 plt.show()
97
98 # Graficar los errores con fechas en el eje X
99 errores = Y_real - Y_pred
100 plt.figure(figsize=(15, 6))
101 plt.plot(fechas_predicciones, errores, color="red", label="
      Errores")
102 plt.title("Errores entre Prediccion y Valores Reales")
103 plt.xlabel("Fecha")
104 plt.ylabel("Error (MW)")
105 plt.legend()
106 plt.grid()
107 plt.gca().xaxis.set_major_locator(mdates.AutoDateLocator())
108 plt.gca().xaxis.set_major_formatter(mdates.
      ConciseDateFormatter(mdates.AutoDateLocator()))
109 plt.xticks(rotation=45)
110 plt.tight_layout()
111 plt.show()

```

Anexo 10: Gráficos usando RStudio

```

1 # Grafico con ggplot data completa
2
3 ggplot(data, aes(x = FECHA, y = EJECUTADO)) +
4 geom_line(color = "blue") +
5 labs(title = "Demanda de Energia Electrica",
6 x = "Fecha", y = "Demanda (MW)") +
7 theme_minimal()

```

```

8
9 # Filtrar solo el 14 de marzo de 2018
10 subset_data <- subset(data, FECHA >= as.POSIXct("2018-07-14
    ") & FECHA < as.POSIXct("2018-07-15"))
11
12 # Cargar ggplot2
13 library(ggplot2)
14
15 # Graficar con linea azul, hora + fecha en el eje X
16 ggplot(subset_data, aes(x = FECHA, y = EJECUTADO)) +
17 geom_line(color = "blue") +
18 labs(
19 x = "Hora y fecha", y = "Demanda (MW)") +
20 scale_x_datetime(date_labels = "%H:%M %d-%m", date_breaks =
    "2 hours") +
21 theme_minimal() +
22 theme(axis.text.x = element_text(angle = 45, hjust = 1))
23
24
25 # Filtrar 2 semanas: del 14 al 28 de marzo de 2018
26 subset_data <- subset(data, FECHA >= as.POSIXct("2018-03-14
    ") & FECHA < as.POSIXct("2018-03-28"))
27
28 # Cargar ggplot2
29 library(ggplot2)
30
31 ggplot(subset_data, aes(x = FECHA, y = EJECUTADO)) +
32 geom_line(color = "blue") +
33 labs(x = "Fecha", y = "Demanda (MW)") +

```

```

34 scale_x_datetime(date_labels = "%d-%m-%Y", date_breaks = "1
      day") +
35 theme_minimal() +
36 theme(axis.text.x = element_text(angle = 45, hjust = 1))
37
38 # Filtrar solo el año 2019
39 data_2019 <- subset(data, format(FECHA, "%Y") == "2019")
40
41 # Graficar
42 ggplot(data_2019, aes(x = FECHA, y = EJECUTADO)) +
43 geom_line(color = "blue") +
44 labs(title = "Demanda de Energia en 2019",
45 x = "Fecha", y = "Demanda (MW)") +
46 scale_x_datetime(date_labels = "%d-%m", date_breaks = "1
      month") +
47 theme_minimal() +
48 theme(axis.text.x = element_text(angle = 45, hjust = 1))

```