

Universidad Nacional de Ingeniería

Facultad de Ingeniería Mecánica



TESIS

Prognosis de la Generación de Potencia del SEIN para reducir costos de facturación de los Usuarios Libres

Para obtener el título profesional de Ingeniero Mecatrónico

Elaborado por

Erick Del Piero Y Daniel Gonzales Requejo

 [0000-0001-6189-304X](https://orcid.org/0000-0001-6189-304X)

Asesor

MSc. Iván Arturo Calle Flores

 [0000-0001-8358-7836](https://orcid.org/0000-0001-8358-7836)

LIMA – PERÚ

2025

Citar/How to cite	Gonzales Requejo [1]
Referencia/Reference	[1] E. Gonzales Requejo, “ <i>Prognosis de la Generación de Potencia del SEIN para reducir costos de facturación de los Usuarios Libres</i> ” [Tesis]. Lima (Perú): Universidad Nacional de Ingeniería, 2025.
Estilo/Style: IEEE (2020)	

Citar/How to cite	(Gonzales, 2025)
Referencia/Reference	Gonzales, E. (2025). <i>Prognosis de la Generación de Potencia del SEIN para reducir costos de facturación de los Usuarios Libres</i> . [Tesis, Universidad Nacional de Ingeniería]. Repositorio institucional Cybertesis UNI.
Estilo/Style: APA (7ma ed.)	

Dedicatoria

Dedico este trabajo a mi familia, especialmente a mi madre, quien me enseñó a soñar y luchar por cumplir mis objetivos, mi abuela que siempre ha sido mi fiel compañera, mi hermana por siempre tener un amor incondicional conmigo y mis sobrinos a quienes amo con todo mi ser.

Agradecimientos

Agradezco a cada uno de mis maestros, que contribuyeron a mi crecimiento intelectual a lo largo de toda mi etapa educativa. De igual manera a las empresas que permitieron adquirir la experiencia necesaria para mi desarrollo profesional, logrando así despertar mi curiosidad e interés por este tema, que espero aporte a mejores desarrollos en el país.

Resumen

La prognosis de la Máxima Demanda Coincidente (MDC) en el Sistema Eléctrico Interconectado Nacional (SEIN) es fundamental para la optimización de costos de los Usuarios Libres (UL). Esta investigación desarrolla un modelo basado en redes neuronales de memoria a largo plazo (LSTM) para la prognosis de la generación de potencia, con el fin de anticipar los momentos en los que ocurre la MDC y facilitar la implementación de estrategias de eficiencia como peak shaving y load shifting.

El estudio analiza datos históricos del SEIN, considerando variables endógenas y exógenas como feriados y días de la semana, para evaluar su impacto en la demanda. Se probaron diversas arquitecturas de modelos predictivos y ventanas de entrada, utilizando métricas como MAE , $RMSE$, $MAPE$ y R^2 para determinar el modelo más robusto. Los resultados muestran que el 92.7% de las MDC ocurren entre las 18:45 h y las 20:45 h, lo que permite reducir el rango de aplicación de estrategias de optimización de la demanda. Además, en un escenario de tres días en la predicción semanal, el modelo logró predecir correctamente la MDC en 10 de los 12 meses evaluados.

El trabajo demuestra la viabilidad de aplicar modelos avanzados de machine learning para mejorar la gestión de la demanda eléctrica, optimizando los costos energéticos de los UL. Se recomienda explorar futuras mejoras mediante la incorporación de nuevas variables exógenas y la integración del modelo con sistemas de almacenamiento energético para una automatización efectiva de las estrategias de gestión de la demanda.

Palabras clave: Prognosis de generación, redes neuronales LSTM, optimización energética, SEIN, peak shaving, load shifting.

Abstract

The accurate prediction of the Maximum Coincident Demand (MDC) in the Peruvian National Interconnected Electric System (SEIN) is crucial for optimizing energy management and reducing transmission and power tariff costs for free users (UL). This research proposes a predictive model based on Long Short-Term Memory (LSTM) neural networks to forecast power generation with a 15-minute granularity. The study evaluates different model architectures and incorporates exogenous variables, such as holidays and weekdays, to enhance prediction accuracy.

The results demonstrate that machine learning techniques, particularly LSTM models, effectively identify patterns in power generation, allowing for a more precise prognosis of MDC occurrences. The findings indicate that MDC predominantly occurs between 18:45h and 20:45h, aligning with demand peaks. The study achieves a 10 out of 12-month prediction accuracy when implementing a strategy based on three weekly MDC probability days. Furthermore, the integration of MDC prediction into energy management strategies, such as peak shaving and load shifting, enables more efficient demand planning, reducing costs and improving grid stability.

Although the model exhibits robust performance, challenges remain in accurately forecasting demand fluctuations during holidays and capturing peak and valley values. Future research should explore additional exogenous variables and hybrid approaches, such as convolutional neural networks (CNNs), to enhance model generalization and improve forecasting accuracy.

Keywords — Power generation forecasting, LSTM, peak shaving, load shifting, SEIN.

Tabla de Contenido

Resumen	v
Abstract	vi
Introducción	xvi
Capítulo I. Generalidades	1
1.1 Descripción del problema de investigación	2
1.2 Objetivos del estudio	4
1.2.1 Objetivo general	4
1.2.2 Objetivos específicos	4
1.3 Hipótesis.....	4
1.3.1 Hipótesis General	5
1.3.2 Hipótesis Específicas	5
1.4 Antecedentes investigativos	5
1.5 Diseño de la Investigación	9
1.5.1 Enfoque Metodológico	9
1.5.2 Operacionalización de Variables	10
Capítulo II. Marco teórico y conceptual	12
2.1 Marco teórico	12
2.1.1 Machine Learning (ML)	12
2.1.2 Series Temporales	14
2.1.2.1 Descomposición Aditiva	15
2.1.2.2 Descomposición Multiplicativa	15
2.1.3 Modelos de Predicción de Series Temporales	16
2.1.3.1 SARIMA	16
2.1.3.2 Prophet	17
2.1.3.3 Long Short-Term Memory (LSTM)	17
2.1.4 Sistemas de Almacenamiento de Energía (SAE)	18

2.1.5	El Comité de Operación Económica del Sistema Interconectado Nacional	18
2.1.6	El Organismo Supervisor de la Inversión en Energía y Minería	20
2.1.7	Sistema Eléctrico Interconectado Nacional (SEIN)	21
2.1.8	Intercambios internacionales de energía	23
2.2	Marco conceptual	24
2.2.1	Peak Shaving	24
2.2.2	Load Shifting	24
2.2.3	Usuarios Libres:	25
2.2.4	Usuarios Regulados	26
Capítulo III. Construcción del Modelo Predictivo		27
3.1	Recopilación y Tratamiento de Datos	29
3.1.1	Fuentes de Datos	29
3.1.2	Justificación del Período de Análisis (2017-2024)	30
3.1.3	Formato y Extracción de Datos	31
3.1.4	Procesamiento y limpieza de Datos	32
3.1.4.1	Manejo de Datos Faltantes	33
3.1.4.2	Detección y Tratamiento de Valores Atípicos (Outliers)	33
3.1.5	Análisis Exploratorio de Datos (EDA)	34
3.1.5.1	Análisis de la Variable Objetivo: Generación de Potencia del SEIN	35
3.1.5.2	Análisis de Variables Adicionales	40
3.1.5.3	Análisis de Correlación de Variables.....	58
3.1.5.4	Herramientas Utilizadas para el EDA	60
3.1.5.5	Conclusiones del EDA	61
3.1.6	Decisiones Tomadas en Base al EDA	62
3.2	Desarrollo de Modelos de Predicción	63
3.2.1	SARIMA	63

3.2.2 Prophet	68
3.2.3 Long short-term memory (LSTM)	70
3.3 Rango horario de la MDC	92
Capítulo IV. A nálisis de resultados	94
Conclusiones	100
Recomendaciones	102
Referencias Bibliográficas	104
Anexos	106

Lista de Tablas

Tabla 1:	<i>Operacionalización de las variables del modelo de predicción y evaluación del impacto económico en usuarios libres del SEIN.</i>	11
Tabla 2:	<i>Métricas descriptivas de la generación</i>	37
Tabla 3:	<i>Métricas descriptivas de los distribuidores</i>	45
Tabla 4:	<i>Métricas descriptivas de los Usuarios Libres</i>	50
Tabla 5:	<i>Métricas descriptivas de la Potencia Ejecutada</i>	52
Tabla 6:	<i>Métricas descriptivas de la Potencia Programada.....</i>	54
Tabla 7:	<i>Métricas descriptivas de la Potencia Reprogramada</i>	55
Tabla 8:	<i>Métricas descriptivas de la Potencia de Intercambio Internacional</i>	57
Tabla 9:	<i>Performance de los modelos.</i>	72
Tabla 10:	<i>LSTM con diferentes arquitecturas.</i>	72
Tabla 11:	<i>Loss de los modelos con diferentes arquitecturas LSTM.</i>	75
Tabla 12:	<i>Métricas de error de los modelos con diferentes arquitecturas LSTM.</i>	76
Tabla 13:	<i>Modelo LSTM con diferentes features</i>	76
Tabla 14:	<i>Performance de los modelos con diferentes features.</i>	81
Tabla 15:	<i>Modelo LSTM con diferentes configuraciones de entrada.</i>	82
Tabla 16:	<i>Rendimiento del modelo LSTM según tamaño de ventana de entrada.</i>	86
Tabla 17:	<i>Desempeño de modelos LSTM con distintas combinaciones de variables de entrada</i>	86
Tabla 18:	<i>Métricas del desempeño del Modelo X durante el año 2024.....</i>	90
Tabla 19:	<i>Métricas del performance del Modelo Y durante todo el 2024</i>	91
Tabla 20:	<i>Métricas del performance del Modelo XX durante todo el 2024.....</i>	91
Tabla 21:	<i>Métricas del performance de la predicción del Modelo XX durante todo el 2024</i>	94
Tabla 22:	<i>Fecha y hora del MDC mensual del 2024</i>	96

Tabla 23:	<i>Fecha del MDC mensual del 2024 (2 días semanales)</i>	96
Tabla 24:	<i>Fecha del MDC mensual del 2024 (3 días semanales)</i>	97
Tabla 25:	<i>MDC acertado según la cantidad de días de la semana aplicados</i>	97
Tabla 26:	<i>Costos energéticos detallados de un Usuario Libre</i>	98
Tabla 27:	<i>Tabla comparativa de costos usando y sin usar el modelo de predicción MDC en un usuario libre durante el 2024</i>	99

Lista de Figuras

Figura 1:	<i>Comparación del consumo eléctrico de un Usuario Libre con y sin estrategia de Load Shifting frente a la demanda del SEIN</i>	<i>3</i>
Figura 2:	<i>Clasificación de la Inteligencia Artificial en dominios cognitivos y subdominios de aplicación</i>	<i>13</i>
Figura 3:	<i>Mapa actualizado de la infraestructura del Sistema Eléctrico Interconectado Nacional (SEIN) en Perú</i>	<i>22</i>
Figura 4:	<i>Diagrama unifilar que representa la interconexión eléctrica entre el Sistema Interconectado del Perú y Ecuador</i>	<i>24</i>
Figura 5:	<i>Comparación gráfica entre estrategias de gestión de la demanda: Peak Shaving y Load Shifting</i>	<i>25</i>
Figura 6:	<i>Diagrama de flujo de la secuencia general de la investigación</i>	<i>28</i>
Figura 7:	<i>Comportamiento histórico de la potencia eléctrica generada en el Perú entre 2017 y 2024, medido en intervalos horarios</i>	<i>35</i>
Figura 8:	<i>Distribución de frecuencia de la potencia generada por el SEIN, con curva de densidad ajustada</i>	<i>36</i>
Figura 9:	<i>Estacionalidad semanal de la generación de potencia eléctrica observada en el mes de enero de 2023</i>	<i>38</i>
Figura 10:	<i>Descomposición aditiva de la serie de generación de potencia, mostrando componentes de tendencia, estacionalidad y ruido</i>	<i>38</i>
Figura 11:	<i>Descomposición multiplicativa de la generación de potencia, identificando los componentes observados, tendencia, estacionalidad y aleatoriedad ...</i>	<i>39</i>
Figura 12:	<i>Evolución temporal de la potencia total registrada por los distribuidores eléctricos del SEIN entre 2017 y 2024</i>	<i>41</i>
Figura 13:	<i>.Histograma con curva de densidad ajustada para la potencia registrada por los distribuidores del SEIN</i>	<i>41</i>

Figura 14:	<i>Aplicación de filtro mediante percentiles 5%–95% y mediana local para la corrección de valores atípicos en la serie de potencia de los distribuidores</i>	43
Figura 15:	<i>Segundo filtrado de la potencia de los distribuidores eléctricos mediante el método MAD (desviación absoluta de la mediana)</i>	43
Figura 16:	<i>Potencia registrada por los distribuidores del SEIN entre 2017 y 2024, luego del proceso de filtrado de valores atípicos</i>	44
Figura 17:	<i>Histograma y curva de densidad ajustada de la potencia de los distribuidores eléctricos del SEIN, tras la aplicación de filtros de depuración.....</i>	45
Figura 18:	<i>Evolución temporal de la potencia registrada por los usuarios libres del SEIN entre 2017 y 2024.....</i>	46
Figura 19:	<i>Histograma y curva de densidad de la potencia registrada por los usuarios libres del SEIN luego del preprocesamiento de datos</i>	47
Figura 20:	<i>Aplicación del método de filtrado mediante mediana local (MAD) sobre los datos de potencia de usuarios libres del SEIN</i>	48
Figura 21:	<i>Potencia histórica registrada por los usuarios libres del SEIN entre 2017 y 2024, luego del proceso de filtrado y corrección de datos atípicos</i>	48
Figura 22:	<i>Histograma y curva de densidad de la potencia eléctrica de los usuarios libres del SEIN luego de aplicar filtros estadísticos de limpieza.....</i>	49
Figura 23:	<i>Evolución temporal de la potencia ejecutada, programada, reprogramada y generada en el SEIN entre 2017 y 2024</i>	51
Figura 24:	<i>Evolución de la potencia ejecutada en el Sistema Eléctrico Interconectado Nacional (SEIN) entre 2017 y 2024</i>	52
Figura 25:	<i>Histograma y curva de densidad de la potencia ejecutada registrada en el SEIN durante el periodo 2017–2024.....</i>	53
Figura 26:	<i>Evolución temporal de la potencia programada en el Sistema Eléctrico Interconectado Nacional (SEIN) entre 2017 y 2024</i>	53

Figura 27:	<i>Histograma y curva de densidad de la potencia programada en el SEIN entre 2017 y 2024.....</i>	<i>54</i>
Figura 28:	<i>Evolución temporal de la potencia reprogramada en el SEIN entre 2017 y 2024.....</i>	<i>55</i>
Figura 29:	<i>Histograma y curva de densidad de potencia reprogramada del SEIN entre 2017 y 2024.....</i>	<i>56</i>
Figura 30:	<i>Comportamiento de la potencia de intercambio internacional con Ecuador, registrada entre 2017 y 2024, donde los valores negativos indican exportación desde Perú</i>	<i>57</i>
Figura 31:	<i>Histograma y curva de densidad de la potencia de intercambio internacional con Ecuador, registrados entre 2017 y 2024.....</i>	<i>58</i>
Figura 32:	<i>Matriz de correlación de variables energéticas del SEIN entre 2017 y 2024, utilizando coeficientes de correlación de Pearson.....</i>	<i>59</i>
Figura 33:	<i>Resultados del modelo SARIMA aplicado a la serie de potencia eléctrica, con visualización de coeficientes, AIC/BIC y métricas de error</i>	<i>64</i>
Figura 34:	<i>Comparación semanal entre los datos reales de potencia eléctrica y las predicciones generadas por el modelo SARIMA</i>	<i>66</i>
Figura 35:	<i>Comparación entre los valores reales de potencia eléctrica y las predicciones generadas por el modelo SARIMA sobre el conjunto de validación.....</i>	<i>67</i>
Figura 36:	<i>Resultado del test Ljung-Box aplicado a los residuos del modelo SARIMA para evaluar la autocorrelación</i>	<i>67</i>
Figura 37:	<i>Predicción semanal del modelo Prophet con intervalo de confianza.....</i>	<i>69</i>
Figura 38:	<i>Predicción de largo plazo con el modelo Prophet (validación)</i>	<i>69</i>
Figura 39:	<i>Evolución de la pérdida de entrenamiento y validación del modelo LSTM</i>	<i>71</i>
Figura 40:	<i>Comparación de Predicción y Real (modelo LSTM)</i>	<i>71</i>
Figura 41:	<i>Pérdida de entrenamiento del Modelo B (LSTM).....</i>	<i>73</i>

Figura 42:	<i>Predicción semanal del Modelo B (LSTM)</i>	74
Figura 43:	<i>Pérdida de entrenamiento del Modelo C (LSTM)</i>	74
Figura 44:	<i>Predicción semanal del Modelo C (LSTM)</i>	75
Figura 45:	<i>Pérdida de entrenamiento del Modelo 1 (Generación)</i>	77
Figura 46:	<i>Predicción semanal del Modelo 1 (Generación)</i>	77
Figura 47:	<i>Predicción de validación del Modelo 1 (Generación)</i>	78
Figura 48:	<i>Pérdida de entrenamiento del Modelo 2 (Generación y Ejecutado)</i>	78
Figura 49:	<i>Predicción semanal del Modelo 2 (Generación y Ejecutado)</i>	79
Figura 50:	<i>Predicción de validación del Modelo 2 (Generación y Ejecutado)</i>	79
Figura 51:	<i>Pérdida de entrenamiento del Modelo 3 (Ejecutado)</i>	80
Figura 52:	<i>Predicción semanal del Modelo 3 (Ejecutado)</i>	80
Figura 53:	<i>Predicción de validación del Modelo 3 (Ejecutado)</i>	81
Figura 54:	<i>Pérdida de entrenamiento del Modelo Y (1344)</i>	83
Figura 55:	<i>Predicción semanal del Modelo Y (1344)</i>	83
Figura 56:	<i>Predicción de validación del Modelo Y (1344)</i>	84
Figura 57:	<i>Pérdida de entrenamiento del Modelo Z (2688)</i>	84
Figura 58:	<i>Predicción semanal del Modelo Z (2688)</i>	85
Figura 59:	<i>Predicción de validación del Modelo Z (2688)</i>	85
Figura 60:	<i>Pérdida de entrenamiento del Modelo XX</i>	87
Figura 61:	<i>Predicción semanal del Modelo XX</i>	87
Figura 62:	<i>Predicción de validación del Modelo XX</i>	88
Figura 63:	<i>Pérdida de entrenamiento del Modelo YY</i>	88
Figura 64:	<i>Predicción semanal del Modelo YY</i>	89
Figura 65:	<i>Predicción de validación del Modelo YY</i>	89
Figura 66:	<i>Diagrama del modelo LSTM seleccionado, incluyendo variables de entrada y salida</i>	92
Figura 67:	<i>Frecuencia por hora del día de la Máxima Demanda Coincidente (2017–2024)</i>	93

Introducción

En el contexto energético actual, la gestión eficiente de la demanda eléctrica es un desafío clave para garantizar la estabilidad del sistema y minimizar los costos operativos, tanto para las empresas distribuidoras como para los usuarios libres (UL). En el Sistema Eléctrico Interconectado Nacional (SEIN) del Perú, la Máxima Demanda Coincidente (MDC) representa un factor determinante en la facturación eléctrica, ya que su ocurrencia incide directamente en los costos por potencia y peajes de transmisión. Ante esta problemática, la implementación de modelos predictivos avanzados permite optimizar estrategias como el peak shaving y el load shifting, facilitando la planificación del consumo eléctrico y la reducción de costos.

Esta investigación propone el desarrollo de un modelo basado en redes neuronales LSTM para la predicción de la generación de potencia eléctrica en el SEIN, con el fin de identificar con precisión los momentos en los que ocurre la MDC. Se evalúan diferentes arquitecturas de modelos y el impacto de variables exógenas, como los días de la semana y los feriados, para mejorar la precisión de las predicciones. A través del análisis de series temporales con granularidad de 15 minutos, se busca establecer patrones de comportamiento que permitan anticipar la ocurrencia de la MDC con un alto grado de confiabilidad.

Los resultados obtenidos en este estudio evidencian que los modelos de aprendizaje profundo pueden proporcionar una prognosis efectiva de la generación de potencia, permitiendo así a los usuarios libres implementar estrategias más eficientes de gestión de la demanda. Se identificó que el 92.7% de las MDC ocurren entre las 18:45h y las 20:45h, lo que refuerza la importancia de focalizar los esfuerzos en la predicción del día más probable de ocurrencia. Además, se logró un acierto en 10 de los 12 meses evaluados bajo un esquema de implementación de estrategias en tres días semanales, lo que demuestra la viabilidad del modelo en un entorno real.

A pesar de los avances logrados, el modelo enfrenta desafíos en la predicción de eventos atípicos, como feriados y períodos de alta variabilidad en la demanda. Asimismo, el estudio resalta la importancia de explorar enfoques complementarios, como el uso de redes neuronales convolucionales (CNNs) y la incorporación de nuevas variables exógenas para mejorar la precisión del pronóstico.

En síntesis, este trabajo no solo aporta un modelo predictivo aplicable a la realidad operativa del SEIN, sino que también establece una base para futuras investigaciones orientadas a la optimización de la planificación energética. Los hallazgos de esta tesis representan un avance significativo en la integración de herramientas de inteligencia artificial en la gestión de la demanda eléctrica, con el potencial de reducir costos, mejorar la eficiencia energética y fortalecer la estabilidad del sistema eléctrico nacional.

Capítulo I. Generalidades

En el contexto del Sistema Eléctrico Interconectado Nacional (SEIN), la facturación de los usuarios libres (UL) está determinada, en parte, por los costos asociados a potencia y peajes de transmisión, los cuales se calculan en función del consumo registrado durante la Máxima Demanda Coincidente (MDC) del mes. La MDC corresponde al momento dentro de la hora punta (17:00h - 23:00h) de lunes a sábado, en el cual el SEIN alcanza su mayor demanda del mes. Dado que este evento depende del consumo agregado de todo el sistema y no exclusivamente del consumo individual del usuario, su determinación precisa resulta un desafío para la optimización de costos energéticos.

Los usuarios libres, al desconocer con antelación los momentos en los que ocurrirá la MDC, enfrentan dificultades para implementar estrategias de gestión del consumo que les permitan reducir su exposición a altos costos de facturación. Estrategias como el Peak Shaving, que consiste en reducir la demanda de la red de manera que no sobrepase un watermark, mediante el uso de almacenamiento energético o fuentes alternativas, y el Load Shifting, que implica desplazar el consumo hacia períodos de menor carga, solo pueden aplicarse de manera efectiva si se dispone de información confiable sobre la probabilidad de ocurrencia de la MDC. La falta de una herramienta predictiva que proporcione esta información con precisión genera sobrecostos en la facturación mensual de los UL, afectando su competitividad y eficiencia operativa.

Desde una perspectiva macroeconómica, la predicción de la MDC no solo beneficia a los usuarios libres en términos de costos, sino que también contribuye a la estabilidad del SEIN. La posibilidad de gestionar de manera eficiente los picos de demanda reduce el estrés en la infraestructura de generación y transmisión, optimizando el uso de los recursos energéticos y favoreciendo un suministro eléctrico más estable. En este sentido, el desarrollo de modelos predictivos que permitan anticipar la ocurrencia de la MDC constituye una herramienta fundamental para la planificación operativa de los grandes consumidores de energía y para la eficiencia global del sistema eléctrico.

Dado que la variable a predecir corresponde a una serie temporal, el estudio explorará diferentes enfoques metodológicos para el desarrollo de un modelo predictivo adecuado. Inicialmente, se realizará un análisis estadístico descriptivo de los datos históricos del SEIN con el fin de identificar patrones de comportamiento en la demanda. Posteriormente, se evaluarán modelos tradicionales de series temporales como SARIMA, modelos basados en descomposición de tendencia y estacionalidad como Prophet de Meta, y modelos de aprendizaje profundo como redes neuronales LSTM, con el propósito de seleccionar el enfoque que ofrezca la mayor precisión en la prognosis de la generación del sistema y, por ende, en la predicción de la MDC.

1.1 Descripción del problema de investigación

En el Sistema Eléctrico Interconectado Nacional (SEIN), los usuarios libres (UL) tienen un esquema de facturación distinto al de los usuarios regulados, ya que pueden negociar directamente con generadores o distribuidores los costos asociados a su consumo eléctrico. Dentro de esta facturación, dos conceptos clave impactan significativamente sus costos: potencia y peajes de transmisión, ambos determinados en función del consumo registrado durante la Máxima Demanda Coincidente (MDC).

La MDC corresponde al momento dentro de la hora punta (17:00h - 23:00h), de lunes a sábado, en el cual el SEIN alcanza su mayor demanda del mes. En este instante, el consumo de cada usuario libre se registra y se utiliza como base para el cálculo de los costos de potencia y peajes de transmisión, independientemente de si su consumo en otros momentos fue mayor o menor, como se muestra en la figura 1. Esto genera una situación de incertidumbre para los UL, quienes no pueden anticipar con precisión el momento exacto en que se producirá la MDC y, por lo tanto, no pueden aplicar estrategias de optimización de consumo de manera efectiva.

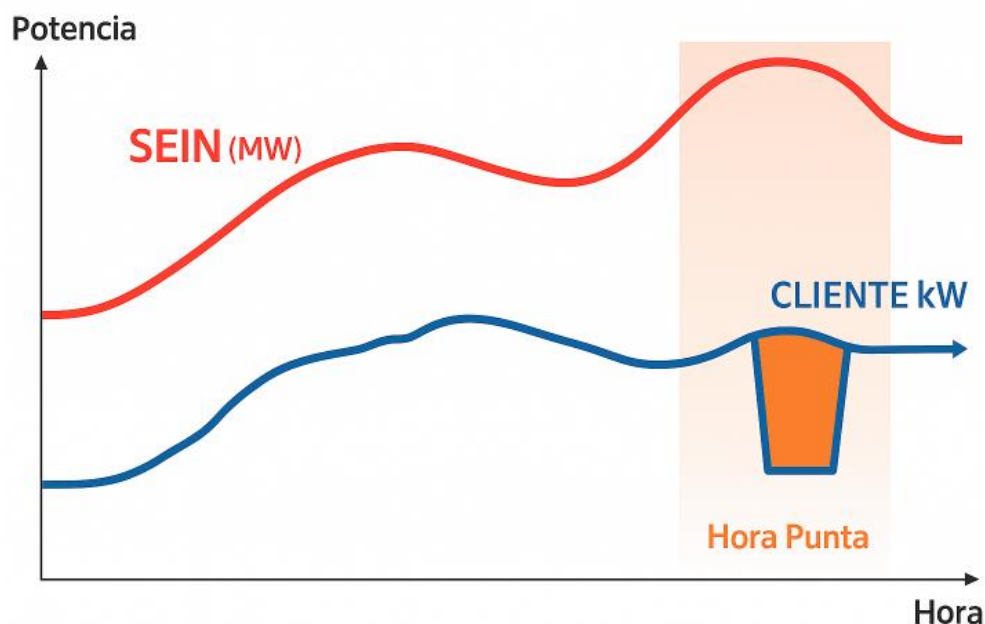
La ausencia de una solución predictiva confiable genera costos innecesarios en la facturación de los UL, quienes, debido a su alto consumo energético, pueden enfrentar facturaciones mensuales que ascienden a cientos de miles de dólares. Por ende, la

capacidad de prever con precisión los momentos de mayor probabilidad de ocurrencia de la MDC representa una oportunidad estratégica para optimizar el consumo y mejorar la eficiencia energética. Con una herramienta predictiva adecuada, los usuarios libres podrían implementar estrategias como el Peak Shaving o el Load Shifting, reduciendo su consumo en los periodos críticos sin comprometer la estabilidad operativa de sus cargas. En el caso de cargas no críticas, los UL podrían reducir su demanda u omitirla temporalmente, mientras que en el caso de cargas críticas podrían recurrir a fuentes de generación alternativas o almacenamiento energético.

Desarrollar un modelo predictivo que permita anticipar con precisión la MDC no solo beneficiaría a los usuarios libres al reducir sus costos de facturación, sino que también contribuiría a la estabilidad del SEIN al mitigar picos de demanda. Este estudio busca abordar este problema mediante la exploración de diferentes modelos de pronóstico de generación, evaluando su efectividad y aplicabilidad en un entorno real para dotar a los usuarios libres de una herramienta predictiva que les permita optimizar su consumo y mejorar su competitividad económica.

Figura 1

Comparación del consumo eléctrico de un Usuario Libre con y sin estrategia de Load Shifting frente a la demanda del SEIN.



Nota: Elaboración propia simulando el comportamiento de carga horaria del usuario libre (kW) y demanda nacional del SEIN (MW).

1.2 Objetivos del estudio

La presente investigación tiene como propósito desarrollar una herramienta predictiva que permita a los usuarios libres (UL) anticipar los momentos de ocurrencia de la Máxima Demanda Coincidente (MDC) dentro del Sistema Eléctrico Interconectado Nacional (SEIN), con el fin de optimizar su consumo energético y reducir los costos asociados a potencia y peajes de transmisión. Para ello, se empleará un enfoque basado en la prognosis de la generación de potencia del SEIN, explorando diferentes modelos de predicción y validando su precisión mediante datos históricos.

1.2.1 Objetivo general

Desarrollar una herramienta predictiva basada en la prognosis de la generación de potencia del SEIN para identificar los momentos de la MDC y facilitar a los usuarios libres la optimización de sus costos de facturación.

1.2.2 Objetivos específicos

- Diseñar y evaluar modelos predictivos de prognosis de generación para determinar los momentos probables de ocurrencia de la MDC.
- Analizar los patrones históricos de generación del SEIN para identificar tendencias que permitan anticipar los momentos de la MDC.
- Validar la precisión de los modelos predictivos mediante datos históricos y actuales del SEIN.
- Proponer recomendaciones prácticas para la implementación de estrategias de optimización de consumo por parte de los usuarios libres.

1.3 Hipótesis

En el presente estudio se plantea que la implementación de un modelo predictivo permitirá anticipar con alta precisión los momentos de ocurrencia de la Máxima Demanda

Coincidente (MDC) en el Sistema Eléctrico Interconectado Nacional (SEIN). Esta predicción facilitará la toma de decisiones estratégicas para los usuarios libres, optimizando su consumo de energía y reduciendo costos asociados a potencia y peajes de transmisión.

1.3.1 Hipótesis General

El desarrollo de una herramienta predictiva basada en la prognosis de la generación de potencia del SEIN permitirá identificar con precisión los momentos de la MDC, facilitando la implementación de estrategias de optimización del consumo energético por parte de los usuarios libres y contribuyendo a la reducción de sus costos de facturación.

1.3.2 Hipótesis Específicas

- La implementación de modelos de predicción permitirá mejorar la precisión en la prognosis de la generación de potencia del SEIN, proporcionando estimaciones confiables de la MDC.
- El análisis de patrones históricos de generación de potencia del SEIN revelará tendencias recurrentes que podrán ser utilizadas para anticipar los momentos de la MDC con alta exactitud.
- La validación de los modelos predictivos mediante datos históricos y actuales del SEIN confirmará la efectividad de la herramienta desarrollada para identificar la MDC con un margen de error aceptable.
- La utilización de la herramienta predictiva en la planificación del consumo energético permitirá a los usuarios libres reducir costos operativos asociados a la potencia contratada y los peajes de transmisión.

1.4 Antecedentes investigativos

Uno de los estudios más recientes en el ámbito de la predicción del consumo energético es el realizado por Antenor Leva Apaza (2023), en el cual se desarrolló un

modelo computacional basado en redes neuronales para predecir el consumo de energía eléctrica en diversas ciudades representativas del Perú. El modelo fue programado en Matlab, permitiendo la carga de datos históricos de consumo para entrenar la red neuronal y mejorar la precisión de las predicciones. Los resultados obtenidos reflejaron una alta precisión en comparación con los datos reales. Sin embargo, este estudio limitó su análisis a un conjunto de ciudades representativas, sin tomar en cuenta el análisis detallado de la información proporcionada por el Comité de Operación Económica del Sistema (COES), que ofrece datos más precisos sobre la generación y demanda de potencia a nivel nacional. Además, el estudio se enfocó únicamente en la predicción del consumo energético, sin considerar la optimización de estrategias como el Load Shifting, que permitiría a los Usuarios Libres ajustar su demanda en momentos estratégicos para reducir costos (Leva Apaza, 2023).

En su tesis, Jairo Curo Martínez (2022) desarrolló un modelo predictivo de la demanda eléctrica diaria del Sistema Eléctrico Interconectado Nacional (SEIN), con el objetivo de reducir la demanda coincidente de un usuario libre del sector industrial. El modelo se basó en redes neuronales artificiales programadas en Python, utilizando datos proporcionados por el Comité de Operación Económica del Sistema (COES) correspondientes al periodo de 2015 a 2021. Los resultados demostraron una alta precisión, con un error porcentual promedio del 0.993 %, y lograron reducir los costos asociados a la demanda coincidente mediante estrategias de Peak Shaving. Sin embargo, este estudio se centró exclusivamente en predicciones diarias de la demanda, lo que obligaba a los usuarios a recurrir diariamente a fuentes alternativas para reemplazar la demanda de la red, sin considerar intervalos de predicción más amplios, como semanas o meses, que permitirían una planificación más eficiente. Además, no se tomó en cuenta el análisis de los datos correspondientes al año 2020, considerados atípicos debido a la pandemia de Covid-19, aspecto que brinda un panorama más completo (Curo Martinez, 2022).

Marvin Alberto Pariona Curi (2022), en su tesis, estudió la estrategia de peak shaving y su influencia en la optimización de los costos de energía en un cliente libre del sector industrial de manufactura. Utilizó un enfoque cuantitativo, con un diseño no experimental, transversal, descriptivo y correlacional, y los resultados demostraron que la implementación de un sistema de autogeneración a gas natural puede reducir significativamente los costos de energía al disminuir la demanda en horas punta. A pesar de que el estudio evaluó de manera efectiva la optimización de los costos a través del peak shaving, se limitó a analizar un cliente libre del sector industrial de manufactura y no incluyó una evaluación más amplia de otros sectores o tipos de clientes libres. Tampoco tomó en cuenta el impacto de la integración de fuentes de energía alternativas en combinación con la autogeneración, aspecto que podría mejorar la eficiencia y reducción de costos en mayor medida (Pariona Curi, 2022).

James Lener Maza Chauca y Miguel Ángel Pérez Félix (2022), en su tesis, investigaron el almacenamiento de energía eléctrica en la central fotovoltaica de Moquegua y su inyección en horas de máxima demanda del Sistema Eléctrico Interconectado Nacional (SEIN). El estudio utilizó un sistema de almacenamiento de energía BESS y evaluó los efectos de inyectar energía durante los momentos de máxima demanda para reducir el uso de centrales termoeléctricas y mitigar el costo de la energía. Los autores concluyeron que la inyección de energía almacenada permitió una significativa reducción en las emisiones de CO₂ y en el costo de generación de energía, con un impacto positivo en la cobertura de la demanda en horas punta. Sin embargo, este estudio no contempló un análisis detallado del comportamiento del Load Shifting ni la optimización de estrategias de reducción de demanda por parte de los usuarios libres, áreas que podrían mejorar la eficiencia del sistema y reducir aún más los costos operativos (Maza Chauca & Pérez Félix, 2022).

Kein Huat Chua, Huoy Lih Bong, Yun Seng Lim, Jianhui Wong y Li Wang (2020) realizaron una revisión exhaustiva sobre las tecnologías de peak shaving más avanzadas y su aplicación en la reducción de la demanda máxima. En su estudio, los autores

clasificaron las tecnologías de peak shaving en cuatro categorías principales: plantas de generación de pico convencionales, sistemas de almacenamiento de energía (ESS), sistemas híbridos de energía renovable y almacenamiento (HREES) y la gestión de la demanda (DSM). La investigación identificó las ventajas y desventajas de cada tecnología, destacando el potencial de los sistemas de almacenamiento de energía para reducir la demanda máxima tanto para las empresas de servicios públicos como para los usuarios finales. Sin embargo, el estudio se centró en la revisión teórica de estas tecnologías sin ofrecer un análisis práctico de su implementación específica en el contexto de redes eléctricas reales. Además, no se profundizó en cómo la integración de tecnologías emergentes como el Internet de las Cosas (IoT) o el análisis de big data podrían optimizar aún más la aplicación de estas tecnologías en la reducción de picos de demanda (Chua, Bong, Lim, Wong, & Wang, 2020).

Ángel Omar Gallo Cruz (2020), en su trabajo de titulación, implementó una herramienta computacional en Python para la minería de datos y la proyección a corto plazo de la demanda horaria de potencia eléctrica del Sistema Eléctrico Interconectado del Ecuador (SNI). El estudio utilizó algoritmos de aprendizaje automático, como Light Gradient Boosting Machine, Random Forest Regressor y Decision Tree Regressor, combinados con el modelo estadístico ARIMA, para generar predicciones más precisas. Uno de los aspectos centrales de su investigación fue la identificación y resolución de problemas en la calidad de los datos, tales como outliers y valores faltantes, los cuales afectaban la precisión de los modelos predictivos. Al corregir estas anomalías, se logró mejorar la fiabilidad de las predicciones a corto plazo. Sin embargo, el enfoque principal del estudio se centró en la mejora de los datos disponibles para aumentar la exactitud de las predicciones, sin explorar la extensión de los horizontes de predicción hacia el mediano o largo plazo, lo que podría proporcionar una visión más amplia para la planificación energética en diferentes escalas temporales (Gallo Cruz, 2020).

1.5 Diseño de la Investigación

1.5.1 Enfoque Metodológico

El estudio adopta un enfoque cuantitativo, dado que su propósito es analizar datos históricos de generación de potencia del SEIN con el fin de desarrollar un modelo predictivo de la Máxima Demanda Coincidente (MDC). La investigación se fundamenta en la recopilación, procesamiento y análisis de datos numéricos mediante técnicas estadísticas y de aprendizaje automático. En este sentido, el enfoque cuantitativo es el más adecuado, ya que permite medir con precisión la relación entre la prognosis de la generación de potencia y la ocurrencia de la MDC, evaluando su impacto económico en los costos de facturación de los Usuarios Libres (UL).

En coherencia con el enfoque adoptado, se emplea un diseño metodológico de tipo no experimental, longitudinal y retrospectivo, dado que no se manipulan las variables del estudio, sino que se analizan datos históricos de generación de potencia del SEIN, registrados en intervalos de 15 minutos durante el periodo 2017–2024. Este diseño permite identificar patrones de comportamiento y tendencias en la generación de potencia que posibiliten anticipar la ocurrencia de la Máxima Demanda Coincidente (MDC). A su vez, se adopta un enfoque descriptivo-correlacional, ya que no solo se describen los patrones observados, sino que también se analiza la relación existente entre dichos patrones y la ocurrencia de la MDC, permitiendo establecer el grado en que las variaciones en la generación pueden ser utilizadas como predictor confiable de los momentos críticos de máxima demanda.

Desde una perspectiva metodológica, el estudio se sustenta en técnicas de modelado de series temporales para el desarrollo de herramientas predictivas, considerando la precisión de los modelos en función de métricas como MAE, MAPE, RMSE y R^2 . Adicionalmente, se realiza una validación con data nueva para evaluar la robustez de los modelos y asegurar su capacidad generalizable a nuevos periodos de tiempo.

1.5.2 Operacionalización de Variables

En la presente investigación se han identificado dos variables clave: la prognosis de la generación de potencia del SEIN como variable independiente y los costos de facturación de los Usuarios Libres (UL) como variable dependiente. La operacionalización de estas variables permite definir cómo serán medidas, evaluadas y analizadas dentro del estudio.

La variable independiente, prognosis de la generación de potencia del SEIN, se define como la predicción de la generación total del sistema eléctrico interconectado nacional en intervalos de 15 minutos. Su objetivo es estimar con antelación los momentos más probables en los que ocurrirá la Máxima Demanda Coincidente (MDC). La evaluación de esta variable se realizará en megavatios (MW), comparando los valores predichos con los valores reales y utilizando métricas de error como el Error Absoluto Medio (MAE), el Error Porcentual Medio Absoluto (MAPE) y la Raíz del Error Cuadrático Medio (RMSE) para determinar la precisión de los modelos predictivos.

Por otro lado, la variable dependiente, costos de facturación de los UL, se define como el impacto que la predicción de la MDC tiene sobre la facturación de los usuarios en lo referente a potencia y peajes de transmisión. No se medirá en términos monetarios exactos, sino mediante la tasa de acierto en la predicción de la MDC, la cual indicará si el modelo logró predecir correctamente los momentos en los que se produjo la máxima demanda que denota la existencia de reducción de costos o no. Para ello, se utilizará un indicador binario donde un valor de 1 representará una predicción acertada (reducción de costos) y un valor de 0 indicará una predicción fallida (no reducción de costos). La tabla 1 muestra la operacionalización de variables.

Tabla 1

Operacionalización de las variables del modelo de predicción y evaluación del impacto económico en usuarios libres del SEIN.

Variable	Definición	Indicadores	Unidad de Medida	Fuente de Datos	Método de Análisis
Prognosis de la generación de potencia del SEIN (V. Independiente)	Predicción de la generación de potencia del SEIN para estimar los momentos de la MDC	MAE, RMSE, MAPE y Generación Pronosticada (MW)	MW (Megavatios), métricas de error	Base de datos del COES: Medidores de Generación e Informe de Evaluación de Operación Diaria (IEOD)	Modelos de series temporales (SARIMA, Prophet, LSTM)
Costos de facturación de los Usuarios Libres (V. Dependiente)	Impacto de la prognosis de la MDC en los costos de potencia y peajes de transmisión.	Tasa de acierto en la predicción de la MDC, indicador binario (1 = acierto, 0 = fallo)	% de predicción correcta	Simulación basada en los datos históricos del SEIN	Evaluación del modelo en la predicción de la MDC

Capítulo II. Marco teórico y conceptual

2.1 Marco teórico

2.1.1 *Machine Learning (ML)*

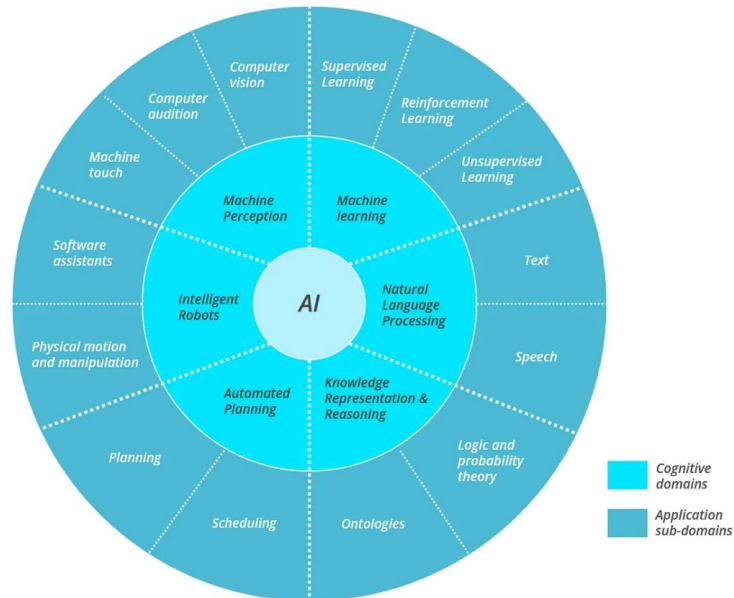
Es una rama de la inteligencia artificial que se enfoca en la creación de sistemas capaces de aprender y mejorar automáticamente a partir de la experiencia, sin ser explícitamente programados. Estos sistemas utilizan algoritmos que identifican patrones en los datos y luego hacen predicciones o toman decisiones basadas en la información aprendida. A lo largo de los años, el machine learning ha revolucionado diversas industrias, permitiendo optimizar procesos, predecir comportamientos y extraer conocimiento valioso de grandes volúmenes de datos.

Dentro de las diversas categorías de ML, el Aprendizaje Supervisado es uno de los enfoques más comunes y ampliamente utilizados. En este método, los algoritmos de aprendizaje reciben un conjunto de datos etiquetados, es decir, datos donde las respuestas correctas ya son conocidas. El objetivo del modelo es aprender a mapear las entradas a las salidas correctas mediante un proceso de entrenamiento que ajusta los parámetros del algoritmo para minimizar los errores. Esto permite que, una vez entrenado, el modelo sea capaz de hacer predicciones precisas sobre nuevos datos no etiquetados. En la figura 2 se observa desde una perspectiva general la relación entre la Inteligencia Artificial, el Machine Learning y el Aprendizaje Supervisado.

En el Aprendizaje Supervisado, el proceso de entrenamiento implica el uso de dos tipos de variables: las variables independientes o características (features), y las variables dependientes o etiquetas (target). El modelo aprende a relacionar las características con las etiquetas a través de métodos como la regresión o la clasificación, dependiendo de si el problema es de naturaleza continua o discreta.

Figura 2

Clasificación de la Inteligencia Artificial en dominios cognitivos y subdominios de aplicación.



Nota: Reproducido de Apanian, T. (2018). A five-minute guide to artificial intelligence. Medium. <https://medium.com/apanian/a-five-minute-guide-to-artificial-intelligence-c4262be85fd3>

Existen varios algoritmos que se utilizan comúnmente en el aprendizaje supervisado, entre los que destacan los árboles de decisión, máquinas de vectores de soporte (SVM), redes neuronales y los modelos lineales, como la regresión lineal o logística. Estos algoritmos varían en su complejidad y en su capacidad para capturar relaciones no lineales en los datos, siendo seleccionados según las características específicas del problema y los datos disponibles.

Una de las principales ventajas del Aprendizaje Supervisado es su capacidad para hacer predicciones precisas cuando se dispone de una cantidad significativa de datos etiquetados y de buena calidad. Sin embargo, este enfoque también presenta desafíos, como la necesidad de una gran cantidad de datos etiquetados para entrenar los modelos y el riesgo de sobreajuste, donde el modelo se ajusta demasiado a los datos de entrenamiento, afectando su capacidad para generalizar en nuevos escenarios.

2.1.2 Series Temporales

Son una sucesión de datos medidos en determinados momentos y ordenados cronológicamente. El análisis de series temporales es un enfoque clave en el estudio de datos que varían con el tiempo, donde la estructura temporal de las observaciones es fundamental para entender el comportamiento subyacente de los fenómenos analizados. En el contexto del Sistema Eléctrico Interconectado Nacional (SEIN), las series temporales de demanda y generación de potencia son esenciales para prever el comportamiento del sistema eléctrico, optimizar el uso de recursos y garantizar la estabilidad operativa. La generación a lo largo del tiempo, como cualquier serie temporal, se puede descomponer en tres componentes principales: tendencia, estacionalidad, y ciclos.

Tendencia: La tendencia en una serie temporal se refiere al movimiento general de los datos a lo largo del tiempo. Este movimiento puede ser creciente, decreciente o mantenerse constante. La tendencia refleja el comportamiento a largo plazo de la variable bajo estudio, excluyendo fluctuaciones a corto plazo o estacionales. En el contexto del SEIN, la tendencia puede evidenciarse en el crecimiento sostenido de la demanda de energía debido al desarrollo económico, el aumento de la población y expansión industrial.

Estacionalidad: La estacionalidad se refiere a patrones que se repiten a intervalos regulares de tiempo, como horas, días, semanas o años. Estos patrones están fuertemente influenciados por factores externos recurrentes, como las condiciones climáticas, los hábitos de consumo o los ciclos económicos. En el SEIN, la estacionalidad es evidente en los cambios cíclicos de la demanda de energía, que suelen depender de la hora del día, el día de la semana o la estación del año.

Ciclos: Los ciclos son fluctuaciones que ocurren en periodos más largos de tiempo y que, a diferencia de la estacionalidad, no necesariamente tienen una periodicidad fija. Los ciclos están influenciados por factores económicos, sociales o tecnológicos y pueden durar varios años. En el SEIN, los ciclos pueden estar relacionados con cambios en la infraestructura energética, variaciones en la economía o la introducción de nuevas fuentes de generación.

Una de las técnicas fundamentales para el análisis de series temporales es la descomposición, la cual permite desagregar la serie en componentes estructurales que facilitan su interpretación y modelado. Existen principalmente dos enfoques: la descomposición aditiva y la multiplicativa, los cuales se detallan a continuación.

2.1.2.1 Descomposición Aditiva

La descomposición aditiva permite representar una serie temporal como la suma de sus tres componentes: tendencia (T), estacionalidad (S) y ruido o residuo (R), bajo el supuesto de que estos elementos son independientes entre sí y se agregan linealmente. Su expresión matemática es:

$$Y_t = T_t + S_t + R_t$$

Este enfoque es adecuado cuando la magnitud de las variaciones estacionales y del error no cambia significativamente con el nivel de la serie. En el contexto del SEIN, se aplica cuando las fluctuaciones diarias o semanales de la demanda se mantienen constantes a lo largo del tiempo, permitiendo identificar con claridad los patrones recurrentes y su evolución.

2.1.2.2 Descomposición Multiplicativa

En la descomposición multiplicativa, se asume que las componentes interaccionan de manera proporcional, y la serie se expresa como el producto de sus elementos:

$$Y_t = T_t \times S_t \times R_t$$

Este modelo es útil cuando la variación estacional y el ruido aumentan o disminuyen en función del nivel de la serie. En series energéticas del SEIN, este comportamiento suele evidenciarse cuando la demanda crece y, en consecuencia, también se incrementa la amplitud de los ciclos estacionales, reflejando una relación no lineal entre los componentes.

2.1.3 Modelos de Predicción de Series Temporales

La predicción de series temporales es una herramienta esencial para la estimación de fenómenos que presentan comportamiento secuencial y dependencias temporales, como es el caso de la generación de potencia eléctrica en el Sistema Eléctrico Interconectado Nacional (SEIN). En el presente estudio se aplicaron tres enfoques de modelado con amplia aceptación en la literatura: SARIMA, Prophet y Long Short-Term Memory (LSTM), los cuales permiten abordar el problema desde perspectivas estadísticas y de aprendizaje profundo. A continuación, se describe cada uno de ellos destacando sus fundamentos, capacidades predictivas y pertinencia en contextos energéticos.

2.1.3.1 SARIMA

El modelo SARIMA (Seasonal AutoRegressive Integrated Moving Average) es una extensión del modelo ARIMA, diseñado para incorporar patrones estacionales en la serie temporal. Su formulación incluye componentes autorregresivos (AR), de media móvil (MA), diferenciación (I), y términos estacionales (SAR, SMA, SD), expresándose como $SARIMA(p,d,q)(P,D,Q)_s$.

Este modelo asume que la serie debe ser estacionaria, por lo que incluye mecanismos de diferenciación para eliminar tendencias y estacionalidades. Resulta especialmente útil cuando se dispone de datos históricos con comportamiento periódico claramente definido, como ocurre en series de demanda energética con ciclos diarios, semanales o anuales.

Una de sus principales ventajas es su bajo costo computacional y alta interpretabilidad. Sin embargo, presenta limitaciones en la captura de relaciones no lineales o patrones más complejos, además de requerir un análisis riguroso para la correcta identificación de los parámetros.

2.1.3.2 Prophet

Prophet es un modelo de pronóstico desarrollado por Facebook que descompone la serie temporal en tres componentes principales: tendencia, estacionalidad y eventos especiales. Está diseñado para facilitar el ajuste automático de patrones complejos con mínima intervención manual, siendo especialmente útil en series con múltiples estacionalidades y eventos atípicos recurrentes.

El modelo asume una estructura aditiva de la forma:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Donde $g(t)$ representa la tendencia, $s(t)$ la estacionalidad y $h(t)$ los eventos especiales. Prophet ofrece una implementación robusta frente a valores atípicos y permite incorporar conocimiento experto, como el calendario de feriados, lo cual es particularmente relevante para series energéticas.

Su facilidad de implementación y capacidad para ajustar múltiples ciclos hacen de Prophet una alternativa atractiva cuando se desea evitar procedimientos complejos de parametrización.

2.1.3.3 Long Short-Term Memory (LSTM)

Las redes neuronales LSTM son una variante de las redes neuronales recurrentes (RNN) diseñadas para modelar secuencias con dependencias a largo plazo. Su arquitectura se compone de celdas de memoria y compuertas de entrada, olvido y salida, lo que les permite retener información relevante durante múltiples pasos temporales sin degradación de gradientes.

A diferencia de los modelos estadísticos, LSTM no requiere suposiciones de linealidad ni estacionariedad, siendo capaz de aprender directamente de los datos. Esta característica las convierte en herramientas especialmente eficaces para series temporales no lineales, con múltiples influencias exógenas y ruido estructurado, como ocurre en el comportamiento de la demanda energética.

Su principal fortaleza radica en su capacidad de generalización y adaptación a patrones complejos. Sin embargo, requieren un volumen considerable de datos y una adecuada regularización para evitar sobreajuste.

2.1.4 *Sistemas de Almacenamiento de Energía (SAE)*

Son sistemas que juegan un papel fundamental en la modernización y operación eficiente de los sistemas eléctricos a nivel global. Su principal función es almacenar energía en momentos de baja demanda o alta generación y liberarla cuando la demanda aumenta o la generación disminuye, mejorando así la estabilidad y la eficiencia del sistema eléctrico. Estos sistemas han ganado protagonismo en las últimas décadas debido al crecimiento de las energías renovables, cuya intermitencia requiere soluciones efectivas para equilibrar la oferta y demanda de energía. Existen varios tipos de sistemas de almacenamiento, cada uno con características específicas que lo hacen adecuado para diferentes aplicaciones en el ámbito energético. Algunos ejemplos son:

- PHS
- Flywheels
- Almacenamiento Térmico
- BESS
- Supercondensadores.

2.1.5 *El Comité de Operación Económica del Sistema Interconectado Nacional*

Es la entidad encargada de coordinar la operación y planificación del Sistema Eléctrico Interconectado Nacional (SEIN) en el Perú. Su rol es fundamental para asegurar el suministro eficiente, seguro y continuo de energía eléctrica en el país, gestionando tanto la operación de corto plazo como la planificación a largo plazo del sistema eléctrico. Además, el COES tiene como objetivo garantizar que el despacho de generación y la transmisión de energía se realicen de manera económica y eficiente, minimizando los costos de operación y maximizando la confiabilidad del suministro.

El COES es un organismo técnico que integra a las empresas generadoras, transmisoras, distribuidoras y a los grandes usuarios del sistema eléctrico. Esta pluralidad de actores le otorga un carácter participativo, lo que permite la toma de decisiones colegiadas y basadas en criterios técnicos. La normativa que regula su funcionamiento está alineada con las disposiciones de la Ley de Concesiones Eléctricas y otras normas del sector eléctrico peruano, asegurando la imparcialidad y transparencia en la operación del SEIN.

Uno de los aspectos más relevantes de la labor del COES es la programación del despacho económico de carga, que consiste en la asignación óptima de las unidades generadoras disponibles para satisfacer la demanda de energía al menor costo posible, tomando en cuenta las restricciones técnicas y operativas del sistema. Esta planificación es realizada en horizontes de corto, mediano y largo plazo, con el objetivo de mantener un equilibrio entre la oferta y la demanda de energía en todo momento, evitando tanto el déficit como los excesos de generación.

Asimismo, el COES es responsable de la coordinación de las interconexiones internacionales de energía, lo cual incluye la gestión de la importación y exportación de electricidad con el país vecino de Ecuador. Estas interconexiones permiten optimizar el uso de los recursos energéticos a nivel regional y contribuyen a la estabilidad del sistema eléctrico peruano.

Otra función crucial del COES es la planificación de la expansión del SEIN, lo que implica evaluar las necesidades futuras del sistema en términos de generación, transmisión y distribución de energía. Esto incluye la identificación de nuevos proyectos de infraestructura que sean necesarios para garantizar la confiabilidad del sistema ante el crecimiento de la demanda y la incorporación de nuevas fuentes de energía, particularmente las energías renovables.

El COES también juega un rol esencial en la supervisión de los mercados de energía y potencia, asegurando que las transacciones entre los diferentes agentes del sector eléctrico se realicen de manera justa y eficiente, respetando los precios y

regulaciones establecidos. En este sentido, su labor es clave para garantizar la competitividad del mercado eléctrico, evitando prácticas monopólicas o abusivas.

2.1.6 El Organismo Supervisor de la Inversión en Energía y Minería

Es la entidad autónoma responsable de supervisar y regular las actividades del sector energético y minero en el Perú, asegurando que las empresas que operan en estos sectores cumplan con las normas de seguridad, calidad y eficiencia establecidas. En el ámbito del mercado eléctrico peruano, OSINERGMIN desempeña un papel fundamental en la regulación del mercado, garantizando la correcta operación del sistema y la protección de los derechos de los usuarios, así como promoviendo la transparencia y la competencia en el sector.

Una de las principales funciones de OSINERGMIN en el mercado eléctrico es la fiscalización de las actividades de generación, transmisión y distribución de energía. Esta función incluye la supervisión de la calidad del servicio eléctrico ofrecido a los usuarios finales, el cumplimiento de las normas técnicas por parte de las empresas generadoras y transmisoras, y la verificación del adecuado mantenimiento y operación de las instalaciones eléctricas en todo el territorio nacional. Este control es esencial para asegurar que el suministro eléctrico sea confiable, seguro y eficiente, contribuyendo al desarrollo económico y social del país.

Además, OSINERGMIN es responsable de la regulación tarifaria en el mercado eléctrico. Esto implica el cálculo y la revisión de las tarifas que pagan los usuarios regulados por el servicio eléctrico. El organismo emplea una metodología basada en criterios técnicos y económicos, con el objetivo de asegurar que las tarifas reflejen los costos reales de suministro, incentivando al mismo tiempo la inversión en infraestructura y la mejora continua de la eficiencia operativa de las empresas del sector. La regulación tarifaria es crucial para garantizar un acceso equitativo a la energía en todo el país y para proteger a los consumidores de precios abusivos; adicionalmente, OSINERGMIN supervisa el funcionamiento competitivo del mercado eléctrico, promoviendo la eficiencia y evitando prácticas monopólicas.

Otro aspecto clave en la relación de OSINERGMIN con el mercado eléctrico peruano es su rol en la fiscalización de los proyectos de infraestructura eléctrica. El organismo se asegura de que los proyectos de expansión y modernización del sistema eléctrico, tales como líneas de transmisión, subestaciones y plantas de generación, cumplan con las normativas vigentes en términos de seguridad, impacto ambiental y eficiencia energética. Esto es fundamental para garantizar que el crecimiento de la infraestructura eléctrica se realice de manera sostenible y alineada con las necesidades del país.

En el contexto de la incorporación de energías renovables en el SEIN, OSINERGMIN tiene la responsabilidad de supervisar que las empresas generadoras de energía renovable cumplan con las normativas específicas del sector, tales como la regulación de la capacidad instalada y los criterios de conexión a la red eléctrica. Además, el organismo apoya la implementación de políticas que promuevan el desarrollo de fuentes de energía limpia, contribuyendo a la diversificación de la matriz energética del Perú y a la reducción de la dependencia de los combustibles fósiles.

2.1.7 Sistema Eléctrico Interconectado Nacional (SEIN)

Es el conjunto de instalaciones eléctricas que permite la interconexión de diversas centrales generadoras, redes de transmisión y subestaciones en el territorio peruano. Su principal función es garantizar el suministro continuo y eficiente de energía eléctrica a los usuarios finales, asegurando la estabilidad y confiabilidad del sistema.

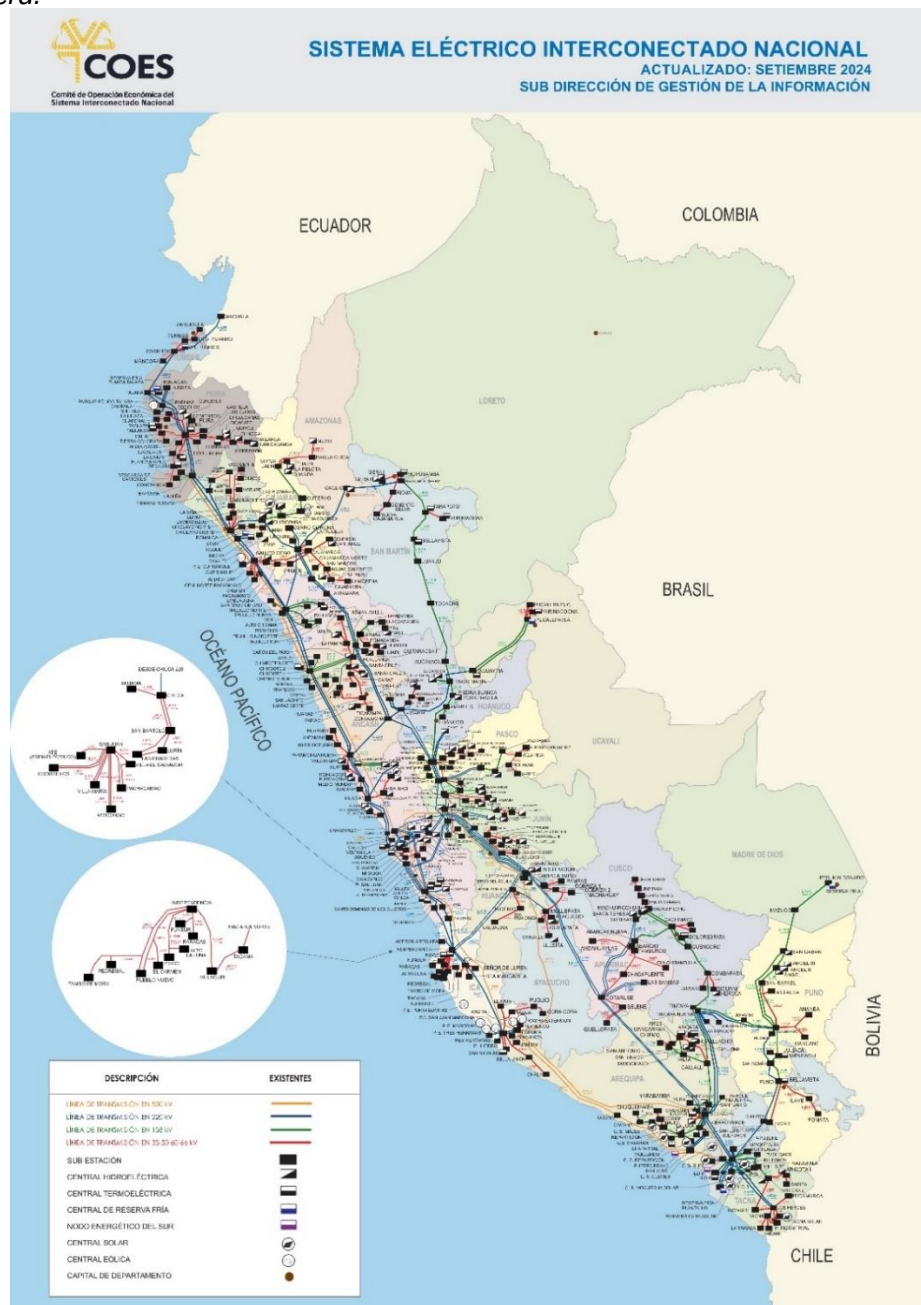
El SEIN está compuesto por varias unidades generadoras de energía, las cuales pueden incluir plantas hidroeléctricas, termoeléctricas, eólicas y solares, que proveen la electricidad necesaria para satisfacer la demanda nacional. Estas unidades están conectadas a una extensa red de transmisión de alta tensión, que distribuye la electricidad a las distintas regiones del Perú, hasta llegar a las redes de distribución locales que suministran energía a los usuarios finales, tanto residenciales, comerciales e industriales.

La operación y gestión del SEIN está a cargo del COES, cuyo objetivo principal es garantizar el uso eficiente de los recursos energéticos y la operación segura del sistema,

coordinando el despacho de las centrales generadoras según la demanda proyectada y las condiciones del mercado. Además, el OSINERGMIN regula y supervisa el funcionamiento del SEIN, velando por el cumplimiento de las normativas y estándares de calidad en la generación, transmisión y distribución de energía. La figura 3 muestra el mapa del SEIN, incluyendo sus redes de transmisión de alto voltaje, subestaciones y tipos de generación.

Figura 3

Mapa actualizado de la infraestructura del Sistema Eléctrico Interconectado Nacional (SEIN) en Perú.



Nota: Reproducido de Comité de Operación Económica del Sistema Interconectado Nacional (COES, 2024),

<https://www.coes.org.pe/Portal/Operacion/CaractSEIN/MapaSEIN>

El SEIN también permite la integración de intercambios internacionales de energía, como es el caso de la interconexión con Ecuador, lo que posibilita la importación y exportación de energía según las necesidades del sistema y las condiciones del mercado. Este sistema de intercambio contribuye a optimizar el flujo de energía dentro de la red y a fortalecer la seguridad energética del país.

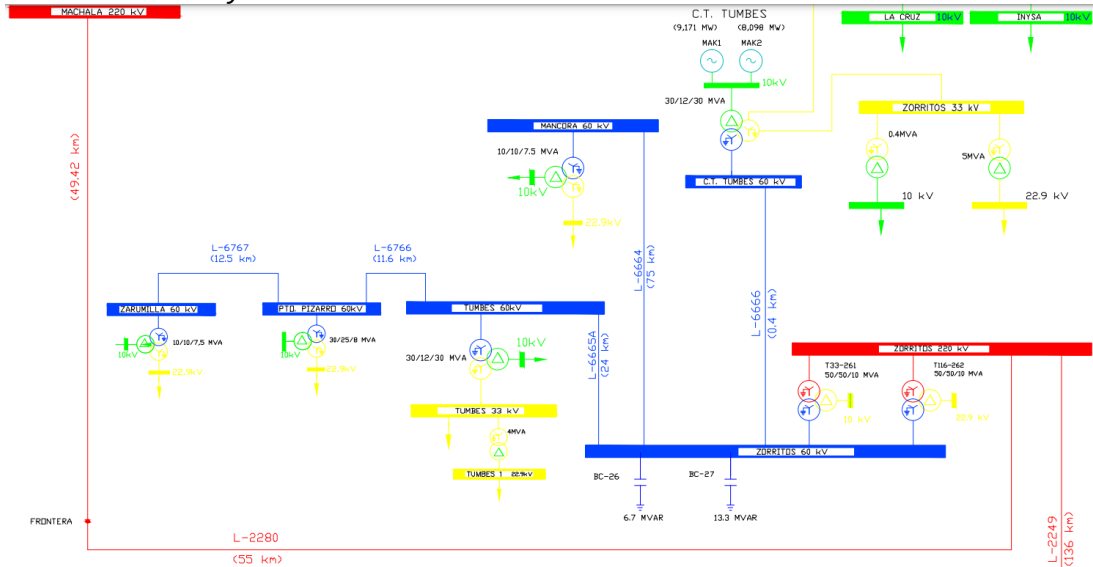
2.1.8 Intercambios internacionales de energía

Son acuerdos bilaterales que permiten la importación y exportación de electricidad entre países, con el fin de optimizar el uso de los recursos energéticos y garantizar la seguridad del suministro eléctrico en situaciones de alta demanda o emergencias. Estos intercambios permiten que los sistemas eléctricos interconectados puedan acceder a fuentes adicionales de generación, equilibrando la oferta y la demanda en momentos críticos, y proporcionando una mayor estabilidad al sistema eléctrico.

En el caso de Perú, el único intercambio internacional de energía vigente se realiza con Ecuador, a través de la línea 2280, que opera a una tensión de 220 kV. Esta línea de transmisión conecta la subestación de Zorritos en Perú con la subestación de Machala en Ecuador, en el diagrama unifilar de la figura 4 se observa dicha interconexión, permitiendo tanto la importación como la exportación de energía eléctrica entre ambos países. El flujo de potencia entre estas dos naciones varía según las condiciones de la red y la demanda energética de cada país, lo que brinda flexibilidad al Sistema Eléctrico Interconectado Nacional (SEIN) para responder ante variaciones en la generación, o el consumo.

Figura 4

Diagrama unifilar que representa la interconexión eléctrica entre el Sistema Interconectado del Perú y Ecuador.



Nota: Reproducido de Comité de Operación Económica del Sistema Interconectado Nacional (COES, 2024), <https://www.coes.org.pe/Portal/Operacion/CaractSEIN/DiagramaUnifilar>

2.2 Marco conceptual

2.2.1 Peak Shaving

Es una estrategia enfocada en la reducción de los picos de demanda eléctrica que superan un umbral previamente establecido, conocido como watermark. En términos prácticos, esta técnica busca "rasurar" los picos de demanda que exceden ese límite, ya que estos picos suelen ser los momentos en los que el costo de la energía es más elevado debido a la necesidad de activar fuentes de generación menos eficientes o más costosas. Al eliminar estos picos, se logra estabilizar la curva de demanda y evitar sobrecostos en el consumo eléctrico, beneficiando tanto a los usuarios como a la red en su conjunto. El Peak Shaving suele implementarse a través de sistemas de almacenamiento de energía, generadores de respaldo o mediante la desconexión temporal de cargas no esenciales.

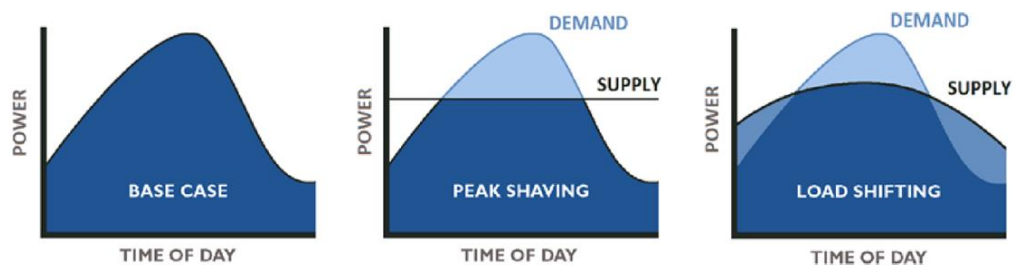
2.2.2 Load Shifting

Se enfoca en el desplazamiento de la demanda eléctrica a momentos específicos del día que resulten más convenientes, desde el punto de vista tanto técnico como

económico. A diferencia del Peak Shaving, que busca reducir picos de demanda, el Load Shifting consiste en trasladar el consumo de energía desde momentos de alta demanda a períodos en los que la energía es más barata o la red tiene mayor capacidad disponible. Esta estrategia no solo implica mover la demanda, sino que también permite reducir la demanda en un punto específico utilizando fuentes diferentes a la red, como sistemas de autogeneración o almacenamiento. De este modo, el Load Shifting ofrece flexibilidad en la gestión del consumo eléctrico y contribuye a mejorar la eficiencia del sistema. La figura 5 muestra el diagrama con una explicación visual de las diferencias entre el Peak Shaving y el Load Shifting.

Figura 5

Comparación gráfica entre estrategias de gestión de la demanda: Peak Shaving y Load Shifting.



Nota: Reproducido de Constellation Energy (2019), <https://www.constellation.com/energy-101/energy-strategy/peak-shaving-vs-load-shifting.html>

2.2.3 Usuarios Libres:

Son aquellos consumidores que tienen la capacidad de negociar directamente los términos de su suministro eléctrico, pudiendo elegir entre adquirir energía de empresas generadoras o distribuidoras. Para ser considerado un Usuario Libre, el consumo de potencia debe ser, como mínimo, de 2.5 MW. Sin embargo, existe una franja opcional que aplica a los consumidores con una demanda de potencia entre 200 kW y 2.5 MW, quienes pueden optar por mantenerse bajo el régimen de Usuarios Regulados o pasar al régimen de Usuarios Libres, dependiendo de sus preferencias y estrategias de consumo. Este poder de decisión otorga a los Usuarios Libres flexibilidad para negociar mejores tarifas y

condiciones de suministro en función de sus necesidades y la dinámica del mercado eléctrico.

2.2.4 Usuarios Regulados

Son aquellos que no cumplen con los requisitos para convertirse en Usuarios Libres o que, teniendo la opción, deciden permanecer bajo el régimen regulado. Estos usuarios están sujetos a las tarifas y condiciones establecidas por el regulador, que en Perú es el OSINERGMIN, el cual define las tarifas basándose en los costos operativos de generación, transmisión y distribución, garantizando así el acceso a la energía a precios justos y regulados.

Capítulo III. Construcción del Modelo Predictivo

Este capítulo presenta el desarrollo detallado del trabajo de investigación, abarcando desde la recopilación y tratamiento de los datos, hasta el modelado, selección, implementación y evaluación del modelo predictivo, así como el análisis de su aplicabilidad y sus limitaciones. Cada una de estas etapas ha sido diseñada para garantizar la eficiencia del proceso y asegurar la elección del modelo más adecuado para cumplir con los objetivos planteados y atender la problemática identificada.

Con la finalidad de proporcionar una visión estructurada del proceso metodológico seguido, la Figura 6 expone un diagrama de flujo que resume la secuencia general de la investigación. Este proceso se organiza en dos grandes fases: la primera corresponde a la recopilación y tratamiento de datos, mientras que la segunda se centra en el desarrollo de los modelos de predicción.

En la fase uno, se inicia con la obtención de información a partir de fuentes oficiales, como el Comité de Operación Económica del Sistema (COES) y el Organismo Supervisor de la Inversión en Energía y Minería (OSINERGMIN), que proporcionan datos estructurados y no estructurados, respectivamente. Posteriormente, se define el período de análisis con base en criterios de relevancia y representatividad, dado el gran volumen de datos disponible. Luego, se lleva a cabo la extracción y estandarización de los datos, uniformando el formato con el que se trabajará en toda la investigación. El procesamiento y limpieza de datos permite asegurar la calidad y utilidad de la información, identificando y corrigiendo posibles inconsistencias o valores atípicos. Finalmente, se realiza un análisis exploratorio de datos, mediante el cual se estudian sus características estadísticas descriptivas, facilitando la toma de decisiones fundamentadas para el modelado.

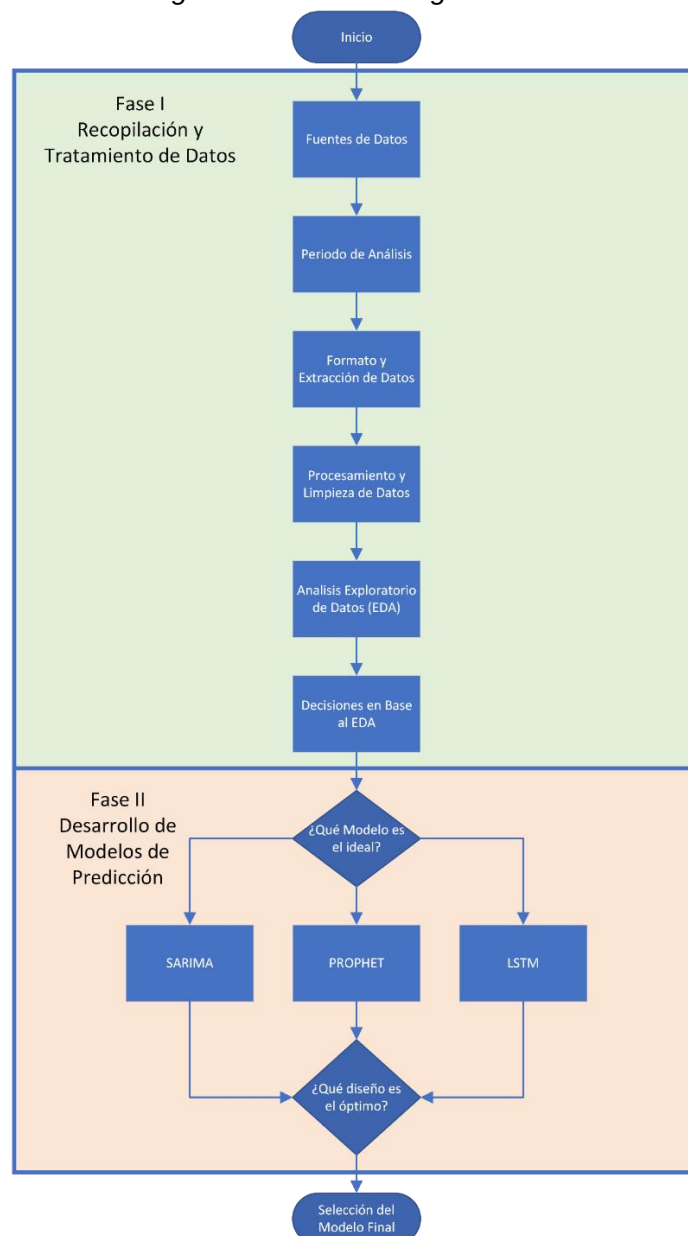
En la fase dos se procede al desarrollo de los modelos predictivos. Se construyen y comparan tres enfoques distintos: SARIMA, Prophet y redes neuronales LSTM. A través de métricas de rendimiento y criterios de ajuste, se selecciona el modelo que mejor se adecúa a los objetivos de la investigación. Una vez identificado el modelo óptimo, se

efectúa un análisis detallado de su arquitectura, evaluando aspectos como el número de capas, el tamaño de las ventanas de entrada, las variables exógenas incorporadas y los parámetros de entrenamiento.

Como resultado de ambas fases, se obtiene un modelo diseñado que responde eficazmente a los requerimientos de la investigación. Finalmente, se evalúan sus predicciones y se analizan sus resultados, los cuales constituyen la base para la elaboración de las conclusiones del estudio.

Figura 6

Diagrama de flujo de la secuencia general de la investigación.



Nota: Elaboración propia usando la herramienta de Microsoft Visio

3.1 Recopilación y Tratamiento de Datos

3.1.1 Fuentes de Datos

La presente investigación se basa en datos obtenidos del Comité de Operación Económica del Sistema Interconectado Nacional (COES), la entidad encargada de coordinar la operación del Sistema Eléctrico Interconectado Nacional (SEIN). Dado que el COES centraliza la recopilación y el procesamiento de información energética en Perú, constituye la fuente oficial más confiable para el análisis de la generación de potencia y la demanda del SEIN. Esta entidad recibe reportes diarios y mensuales de todos los agentes del mercado eléctrico, información que utiliza para la toma de decisiones regulatorias y económicas, particularmente en lo relacionado con la facturación de potencia y peajes de transmisión durante la Máxima Demanda Coincidente (MDC).

A nivel normativo, si bien la regulación del sector eléctrico está bajo la supervisión del Organismo Supervisor de la Inversión en Energía y Minería (OSINERGMIN), el COES es responsable de estructurar y publicar la información operativa del SEIN. Por ello, las bases de datos del COES proporcionan la información más detallada y actualizada sobre la generación, demanda, intercambios internacionales y otros parámetros relevantes del sistema eléctrico peruano.

Para el desarrollo de este estudio, se han utilizado diversas bases de datos del COES que contienen información histórica del SEIN desde intervalos de 15 minutos, lo que permite modelar la variabilidad del sistema con una granularidad adecuada. A continuación, se describen las principales fuentes de datos empleadas:

Medidores de Generación:

- Contiene los registros de generación de potencia de todas las centrales del SEIN en intervalos de 15 minutos.
- Fuente primaria para construir la prognosis de generación de potencia.

Informe de Evaluación de Operación Diaria (IEOD):

- Registra la operación diaria del SEIN, incluyendo variables relevantes para el modelado de la MDC con granularidad de 30 minutos.
- Aporta información sobre la potencia ejecutada por los usuarios, programada para ejecutarse y reprogramada en caso haya un incidente.

Datos de Demanda de Usuarios Libres y Distribuidores:

- Contiene la información de la demanda tanto de los Usuario Libres como la de los Distribuidores, los cuales sumados equivalen a la generación.
- Están en intervalos de 30 minutos y están segmentados por cada usuario y distribuidor en particular.

Registros de Contadores de Energía:

- Muestran el intercambio tanto de importación como de exportación con Ecuador, con granularidad de 15 minutos.
- Permiten evaluar cómo la variabilidad en los flujos energéticos afecta los momentos de máxima demanda del sistema.

3.1.2 Justificación del Período de Análisis (2017-2024)

El período de análisis comprende desde enero de 2017 hasta diciembre de 2024, seleccionado en función de los siguientes criterios:

Disponibilidad y confiabilidad de datos:

- Aunque existen registros de generación desde el año 2000, otras fuentes de datos, como el IEOD, comenzaron a consolidarse a partir de 2017, con inconsistencias en los primeros años.
- A partir de 2017, la información es más precisa, con menos datos faltantes y una estructura más estable.

Representatividad de la infraestructura actual del SEIN:

- Analizar un período demasiado amplio podría no reflejar adecuadamente la evolución reciente del sistema eléctrico peruano.

- En los últimos años, se han producido cambios significativos en la capacidad instalada, en los intercambios internacionales, fuentes renovables y en el comportamiento de la demanda, lo que hace que los datos más recientes sean más relevantes para la predicción de la MDC.

Volumen de datos suficiente para modelos predictivos:

- Se requieren varios años de datos para identificar patrones estacionales y tendencias de largo plazo en la generación y demanda.
- El período 2017-2024 proporciona una cantidad suficiente de información para entrenar y validar modelos de series temporales con robustez estadística.

3.1.3 Formato y Extracción de Datos

Los datos del COES están disponibles en formatos Excel y CSV, lo que facilita su procesamiento mediante herramientas de análisis de datos como Python y R. Durante la fase de exploración y análisis, la descarga de la información se realizó de manera manual, permitiendo evaluar qué variables eran realmente relevantes para el modelo.

Una vez definidas las variables esenciales, se optimizó el proceso mediante scripts automatizados, los cuales filtran y extraen únicamente la información necesaria para el modelo. Aunque en la presente investigación la descarga de datos no se ha automatizado completamente, en futuras implementaciones, se podría incorporar automatización en la descarga de datos desde fuentes oficiales, mejorando así la eficiencia del sistema predictivo.

El uso de datos provenientes del COES garantiza un análisis confiable y representativo del comportamiento del SEIN en los últimos años. La selección de fuentes de información y del período de análisis ha sido realizada con base en criterios de disponibilidad, calidad de datos y representatividad del sistema eléctrico actual. De esta manera, se asegura que los modelos predictivos desarrollados en esta investigación se basen en información precisa y relevante para la estimación de la MDC.

3.1.4 Procesamiento y limpieza de Datos

El preprocesamiento de datos es una etapa imprescindible en la presente investigación, ya que garantiza que la información utilizada en los modelos predictivos sea precisa, coherente y representativa del comportamiento real del Sistema Eléctrico Interconectado Nacional (SEIN). Dado que los datos fueron obtenidos de diferentes fuentes dentro del COES, fue necesario aplicar diversas técnicas para su integración, limpieza y homogeneización antes de su uso en el modelado de la prognosis de generación de potencia. Los datos extraídos del COES presentaban diferentes estructuras y formatos. A continuación, se describen las principales bases de datos utilizadas y sus características:

- **Medidores de Generación:** Se obtuvieron 96 archivos mensuales (2017-2024), cada uno con valores de generación en intervalos de 15 minutos por central generadora. Para construir la serie temporal de la generación total del SEIN, se realizó una sumatoria de todas las centrales en cada punto de muestreo.
- **Informe de Evaluación de Operación Diaria (IEOD):** Se recopilaron 2,922 archivos diarios, cada uno con múltiples hojas, de las cuales se utilizaron tres: potencia ejecutada, potencia programada y potencia reprogramada. Esta información venía en intervalos de 30 minutos, por lo que fue necesario ajustar su granularidad.
- **Demanda de Usuarios Libres (UL) y Distribuidores:** Se extrajeron 96 archivos mensuales con datos en intervalos de 30 minutos y en una estructura de filas y columnas diferente a la de los otros archivos.
- **Registros de Contadores de Energía:** Se utilizaron exclusivamente para analizar cambios en la MDC debido a intercambios internacionales. Estos datos estaban en intervalos de 15 minutos, pero medidos en energía, por lo que se transformaron a potencia equivalente para estandarizar unidades.

Dado que los datos presentaban diferentes granularidades (30 minutos en algunos casos y 15 minutos en otros), se decidió unificar todos los datos en una granularidad de 15 minutos para mantener la coherencia con la variable objetivo. Para ello, se aplicó interpolación lineal, un método matemático sencillo y suficiente para el caso de aplicación.

3.1.4.1 Manejo de Datos Faltantes

Durante la revisión de los datos, se identificó que tres días dentro del período de estudio no contaban con registros en el IEOD debido a una pérdida de información en la base de datos del COES. Los días afectados fueron:

- 31 de julio de 2019
- 4 de noviembre de 2020
- 25 de diciembre de 2020

Para completar estos valores faltantes, se probaron tres métodos de imputación:

- Interpolación con datos de un año anterior y posterior al día faltante.
- Interpolación con datos de una semana antes y una semana después.
- Interpolación con datos del día anterior y el día posterior.

Tras comparar los resultados, se seleccionó la tercera opción (día anterior y posterior), ya que generaba valores más alineados con el contexto inmediato del sistema eléctrico, evitando desfases o anomalías, y respetando la estadística.

3.1.4.2 Detección y Tratamiento de Valores Atípicos (Outliers)

Para identificar valores atípicos en los datos, se aplicaron los siguientes métodos estadísticos:

- Rango Intercuartílico (IQR)
- Z-score (Puntaje Z)
- Percentiles extremos

El análisis de outliers reveló que las variables afectadas fueron las de demanda de UL y distribuidores, la potencia programada y reprogramada. Se observó que los primeros datos contenían anomalías significativas que no reflejaban la realidad del sistema,

afectando la calidad del modelado. Para su corrección, se probaron dos métodos de reemplazo, ambos enfoques fueron evaluados, asegurando que los datos procesados mantuvieran su validez estadística sin distorsionar la información clave. Los métodos de reemplazo son:

- Media móvil para suavizar variaciones abruptas.
- Mediana local para reducir el impacto de valores extremos.

Antes de proceder con el modelado, se verificó que los datos estuvieran correctamente estructurados y listos para su uso. Esto incluyó:

- Homogeneización dimensional: Todas las variables fueron ajustadas a la misma unidad y granularidad temporal (15 minutos).
- Corrección de timestamps: Se aseguró que todos los registros estuvieran alineados temporalmente.

3.1.5 *Análisis Exploratorio de Datos (EDA)*

El Análisis Exploratorio de Datos (EDA) constituye una fase crítica en el proceso de modelado predictivo, ya que permite comprender la distribución y el comportamiento de las variables antes de su uso en la construcción de modelos. Este análisis se enfocó en evaluar la calidad y coherencia de los datos, identificar inconsistencias o valores atípicos, analizar tendencias y patrones en la generación de potencia del SEIN, y validar la relevancia de las variables seleccionadas.

Este análisis facilitó la toma de decisiones clave en la estructuración del dataset, la identificación de variables con un alto grado de ruido o información no representativa, y la selección de aquellas con mayor influencia en la predicción de la Máxima Demanda Coincidente (MDC). Además, permitió corroborar la existencia de patrones estacionales y tendencias históricas en la generación de potencia, aspectos fundamentales para la construcción de un modelo predictivo sólido.

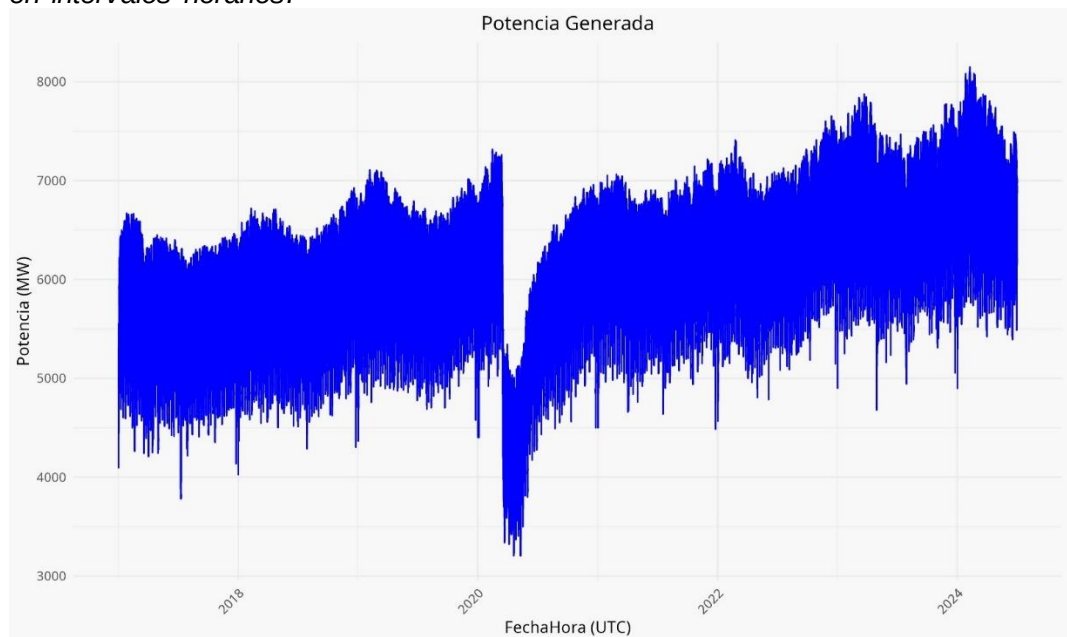
3.1.5.1 Análisis de la Variable Objetivo: Generación de Potencia del SEIN

La generación de potencia en el Sistema Eléctrico Interconectado Nacional (SEIN) representa el total de energía generada por las distintas fuentes de generación en el Perú e inyectada a la red. Esta variable es fundamental en el presente estudio, ya que se considera como el target en la predicción de la Máxima Demanda Coincidente (MDC).

El COES mantiene un registro detallado de la generación de potencia en intervalos de 15 minutos, a través de su base de datos denominada “Medidores de Generación”, en la cual se dispone de información desde el año 2000 hasta la actualidad. En esta base de datos, además de la Potencia Activa, se registran la Potencia Reactiva Inductiva, Potencia Reactiva Capacitiva y Servicios Auxiliares (SS.AA.). Sin embargo, este análisis se centra únicamente en la Potencia Activa, por su relevancia en la determinación de la MDC. A pesar de contar con registros desde el año 2000, se ha delimitado el periodo de análisis entre enero de 2017 y diciembre de 2024 debido a la estabilidad, completitud y representatividad de los datos en ese rango.

Figura 7

Comportamiento histórico de la potencia eléctrica generada en el Perú entre 2017 y 2024, medido en intervalos horarios.

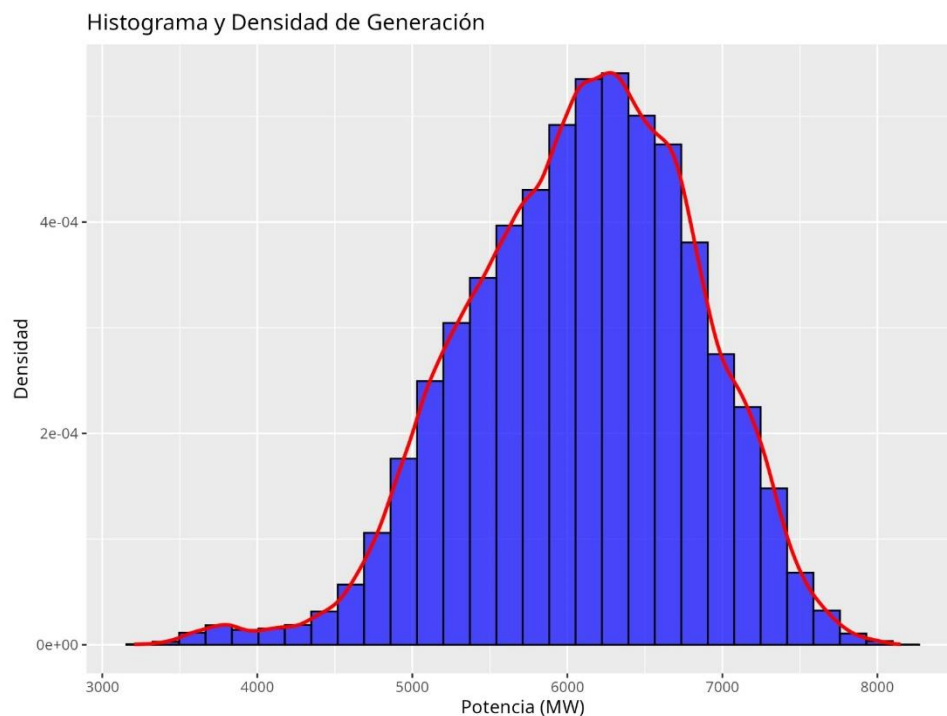


Nota: Elaboración propia en R con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La Figura 7 muestra la evolución de la potencia generada entre 2017 y 2024, donde se observa una caída pronunciada durante el año 2020 debido al impacto del COVID-19. Esta disminución generó valores atípicos en la distribución, lo cual se evidencia en el histograma de la Figura 8, la línea roja es la curva de densidad, que muestra una distribución asimétrica con una ligera cola hacia la izquierda.

Figura 8

Distribución de frecuencia de la potencia generada por el SEIN, con curva de densidad ajustada.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Este comportamiento se corrobora en la Tabla 2, que presenta las principales métricas descriptivas de la generación de potencia:

Tabla 2*Métricas descriptivas de la generación*

Medida	Valor	Unidad	Nota
Media	6088.67	MW	Promedio
Mediana	6135.11	MW	Valor del medio
Varianza	533555.5	MW^2	M. Dispersión
Desviación estándar	730.45	MW	M. Dispersión
Mínimo	3204.64	MW	Registrado durante la Pandemia
Máximo	8149.01	MW	Registrado el último año
Primer cuartil (Q1)	5591.16	MW	-
Tercer cuartil (Q3)	6617.57	MW	-

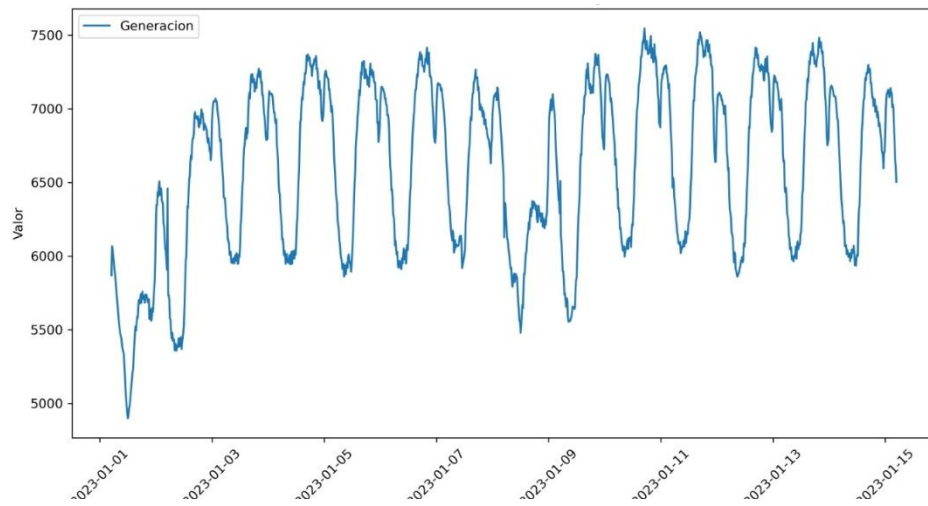
Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Tal como se observa también en la Figura 8, la ligera diferencia entre la media (6088.67 MW) y la mediana (6135.11 MW) confirma una asimetría negativa en la distribución. Además, los valores de desviación estándar y varianza indican una alta variabilidad en la generación, lo que es consistente con la naturaleza fluctuante del consumo eléctrico. En términos de dispersión, el primer y tercer cuartil revelan que la mayoría de los valores de generación se encuentran en el rango de 5591 MW a 6617 MW, lo que representa los valores más frecuentes dentro del SEIN en el período analizado.

Durante la inspección de la data se identificó un patrón semanal que se mantiene de forma consistente a lo largo del período analizado, con el domingo destacando como un día atípico seguido de seis días con comportamiento similar. La figura 9, muestra la representación gráfica de 2 semanas, donde se nota con claridad que inicia un domingo, la figura presenta dos semanas como ejemplo visual.

Figura 9

Estacionalidad semanal de la generación de potencia eléctrica observada en el mes de enero de 2023.

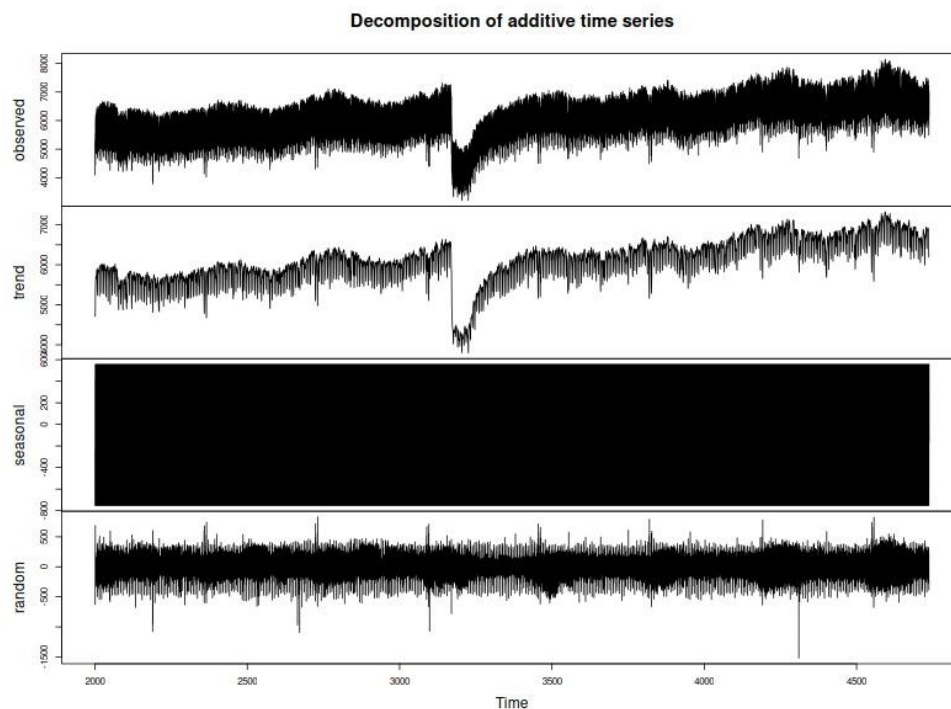


Nota: Elaboración propia en Jupyter Notebook con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Se descompone la serie temporal en forma aditiva y multiplicativa, según lo explicado en los puntos 2.1.2.1 y en 2.2.2.2. Dichas descomposiciones se muestran en la figura 10 y 11, respectivamente. Las cuales muestran la gráfica real observada (observed), la tendencia (trend), la estacionalidad (seasonal) y el ruido (random).

Figura 10

Descomposición aditiva de la serie de generación de potencia, mostrando componentes de tendencia, estacionalidad y ruido.

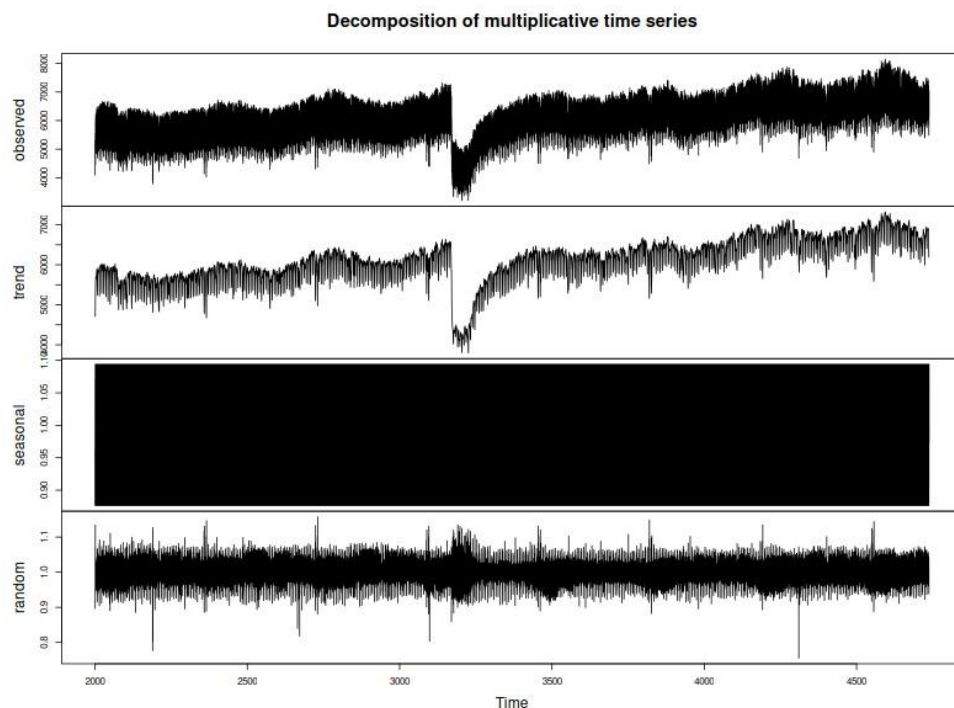


Nota: Elaboración propia en R con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La descomposición aditiva de la serie temporal de generación de potencia del SEIN revela que la tendencia general presenta fluctuaciones de mediano y largo plazo, con una caída abrupta en un periodo específico debido al impacto del COVID-19, lo que constituye un evento atípico en la serie. La componente estacional muestra una variabilidad relativamente constante a lo largo del tiempo, indicando que los patrones de generación siguen una estructura recurrente que no se ve afectada significativamente por la magnitud de la tendencia. Además, el componente residual mantiene una dispersión homogénea, lo que sugiere que los factores aleatorios no explicados por la tendencia y la estacionalidad permanecen relativamente estables en el tiempo.

Figura 11

Descomposición multiplicativa de la generación de potencia, identificando los componentes observados, tendencia, estacionalidad y aleatoriedad.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Por otro lado, la descomposición multiplicativa evidencia que la estacionalidad varía en proporción a la tendencia, es decir, cuando la generación de potencia aumenta, la amplitud de los ciclos estacionales también se incrementa, lo que sugiere una relación no lineal entre estos componentes. La tendencia mantiene un comportamiento similar al modelo aditivo, reflejando eventos estructurales como la pandemia, mientras que la

estacionalidad relativa permite capturar con mayor precisión los cambios cíclicos en función de distintos niveles de demanda.

3.1.5.2 Análisis de Variables Adicionales

Además de la generación de potencia del SEIN, se analizaron variables adicionales provenientes de diferentes fuentes del COES, con el objetivo de evaluar su comportamiento y su posible contribución en la predicción de la Máxima Demanda Coincidente (MDC). Estas variables incluyen:

- Demanda de los Agentes: Usuarios Libres (UL) y Distribuidores.
- Informe de Evaluación de Operación Diaria (IEOD): Potencia ejecutada, programada y reprogramada.
- Registros de Contadores de Energía: Intercambios internacionales de energía.

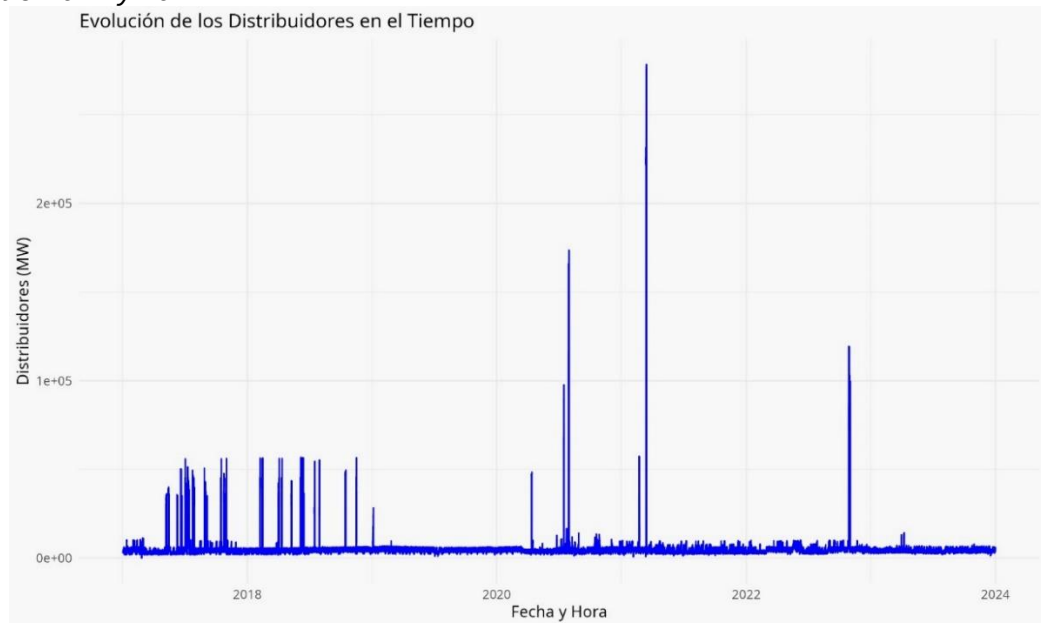
Cada una de estas variables fue analizada mediante gráficos de evolución temporal, histogramas y métricas estadísticas para evaluar su calidad, distribución y relevancia en el estudio.

Distribuidores

La demanda de los distribuidores constituye una parte clave del consumo energético del SEIN, por lo que se analizó su estabilidad y posible aporte en la predicción de la MDC. La información fue extraída de la base de datos del COES y abarca el periodo de enero de 2017 a diciembre de 2023. En la Figura 12, se muestra la evolución de la demanda de los distribuidores a lo largo del tiempo. Se observa que los datos presentan picos extremos que superan valores normales, y son técnicamente imposibles dentro del sistema eléctrico, lo que indica la presencia de outliers severos.

Figura 12

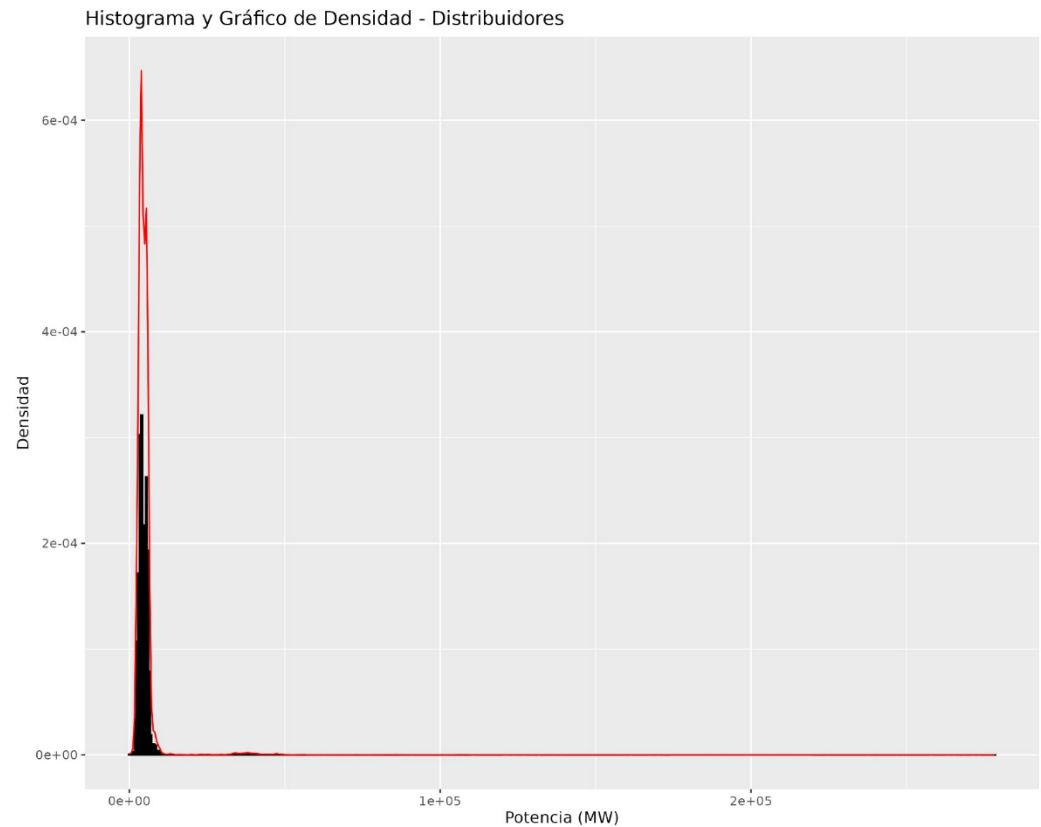
Evolución temporal de la potencia total registrada por los distribuidores eléctricos del SEIN entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 13

Histograma con curva de densidad ajustada para la potencia registrada por los distribuidores del SEIN.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

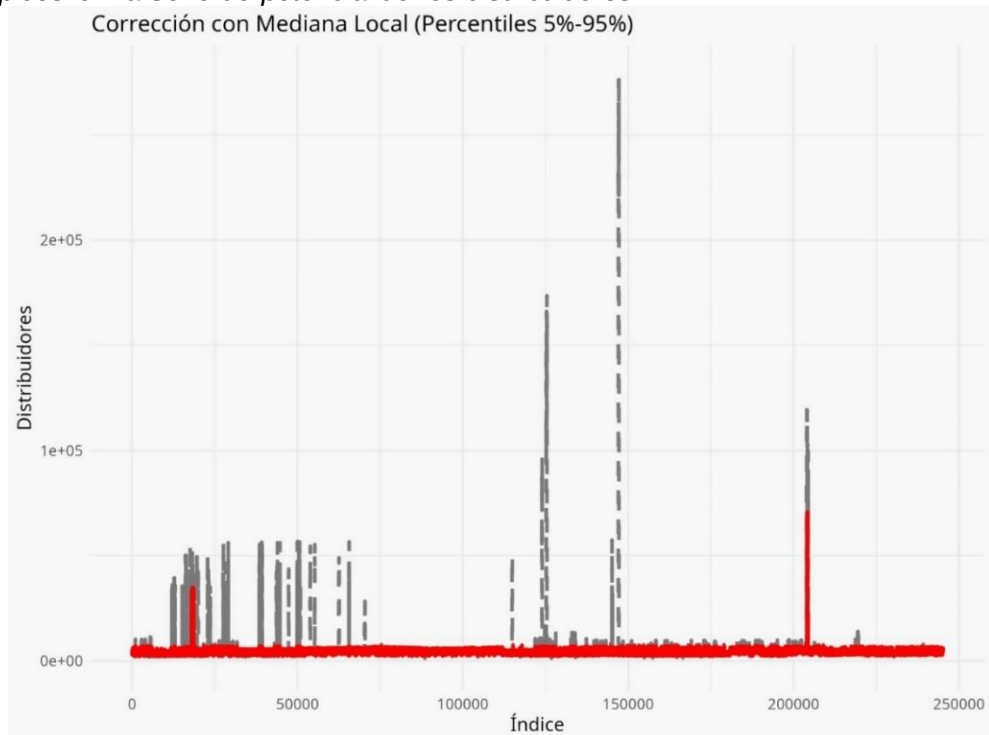
El histograma de la demanda de los distribuidores, presentado en la Figura 13, evidencia la curva de densidad de color rojo con una fuerte asimetría positiva, con una concentración de valores en los niveles bajos y una cola larga hacia la derecha, reflejando la presencia de valores atípicos en el dataset original.

Para mitigar este problema, se procede a hacer una limpieza de datos. Esta limpieza consta de 3 fases:

- Un primer filtrado y reemplazo de outliers, se probaron 3 métodos de detección de outliers: percentiles, MAD y Z_Score. Y para su reemplazo se usó la media móvil y la mediana local. Finalmente, el mejor resultado fue la combinación de percentiles con mediana local, como se muestra en la figura 14 comparando su valor filtrado de color rojo, con sus valores iniciales de color gris sombreado.
- Posteriormente se repite el proceso anterior, y esta vez la mejor combinación se genera por la detección con MAD y reemplazo por la mediana local. En la figura 15 se visualiza el valor filtrado de color rojo y el valor limpiado en forma de sombras negras.
- Finalmente se hace un trabajo manual ya que solo quedaban dos semanas con valores picos. De esta manera se logra el resultado de la figura 16, que es la variable con la que se trabaja.

Figura 14

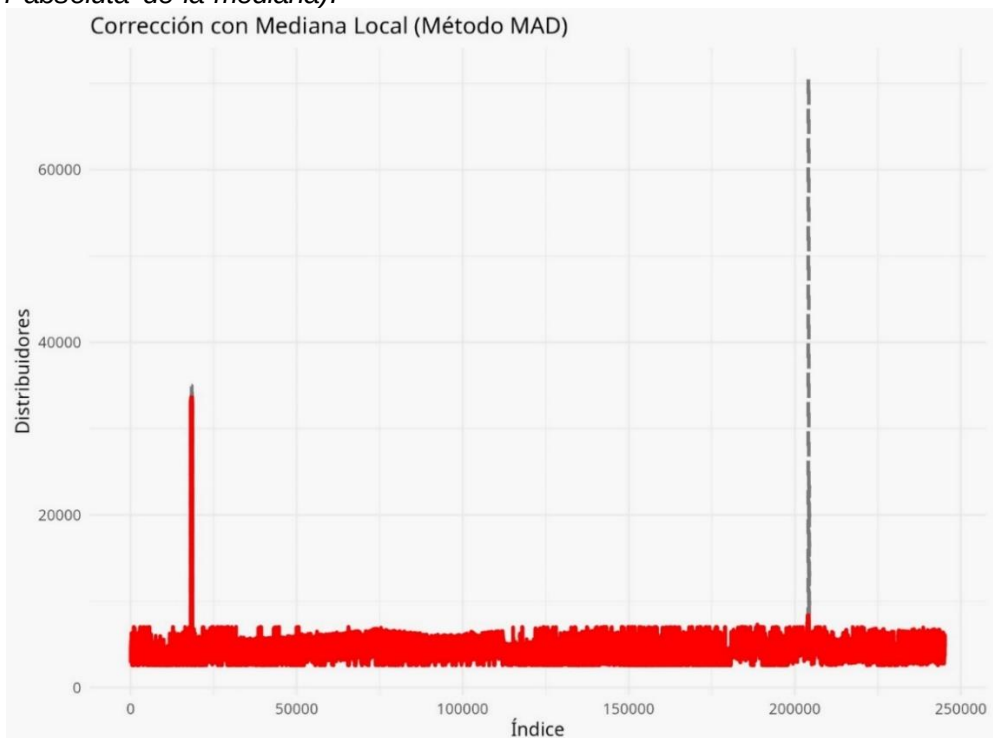
Aplicación de filtro mediante percentiles 5%–95% y mediana local para la corrección de valores atípicos en la serie de potencia de los distribuidores.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 15

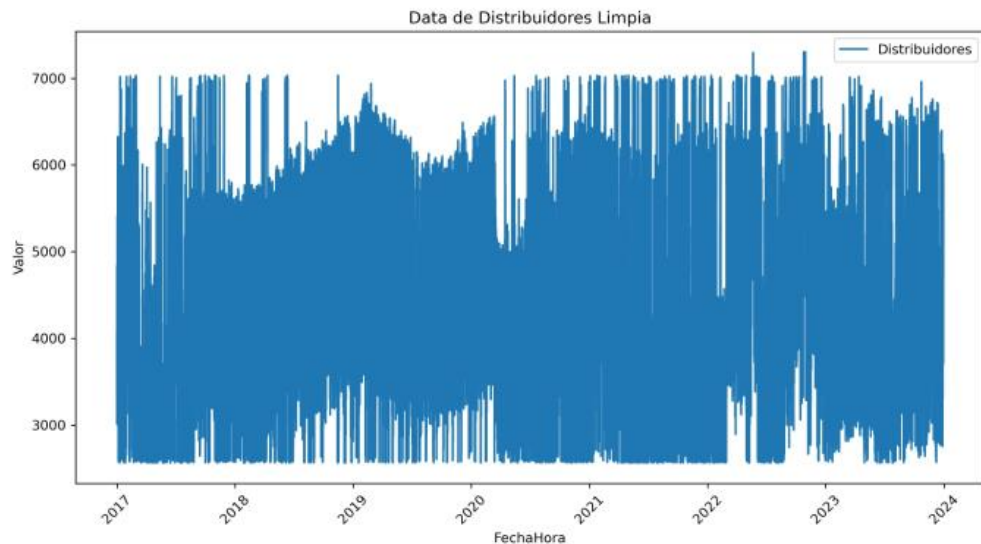
Segundo filtrado de la potencia de los distribuidores eléctricos mediante el método MAD (desviación absoluta de la mediana).



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 16

Potencia registrada por los distribuidores del SEIN entre 2017 y 2024, luego del proceso de filtrado de valores atípicos.



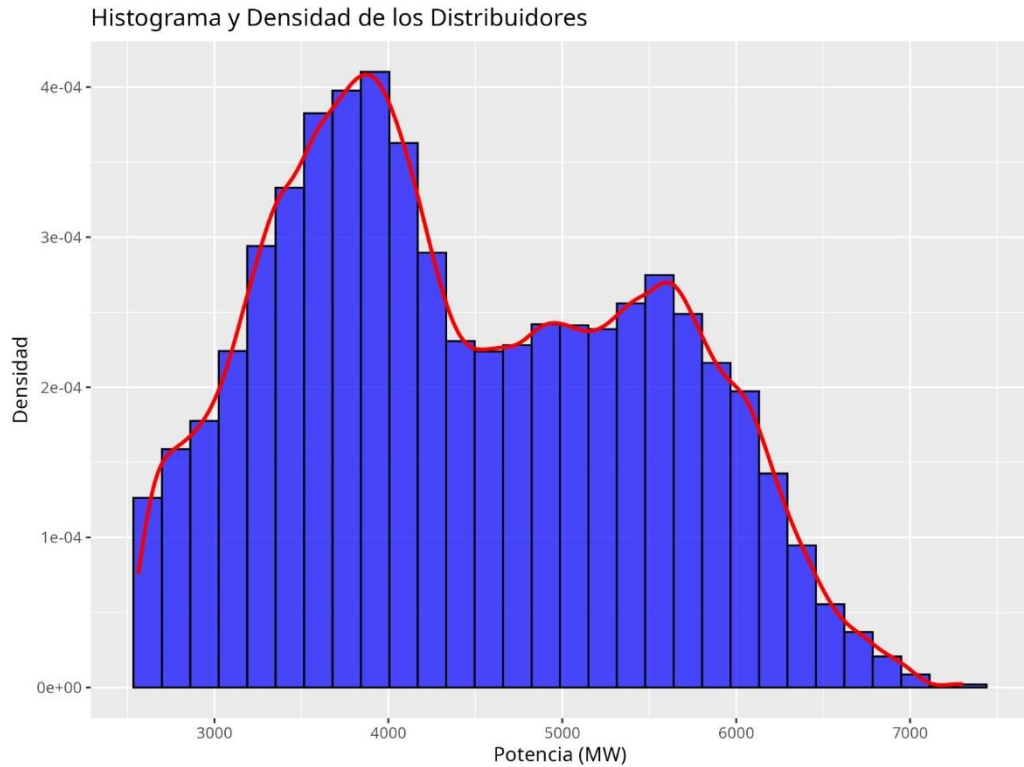
Nota: Elaboración propia en Jupyter Notebook con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Con esta data limpia, se vuelve a graficar el histograma, en la figura 17, donde se muestra la distribución de la demanda de los distribuidores en términos de potencia (MW). Se observa una curva de densidad con una distribución multimodal con al menos dos picos prominentes alrededor de los 4000 MW y 6000 MW. La curva de densidad (en rojo) suaviza la distribución y resalta estas concentraciones de datos.

En la tabla 3 se muestra sus características descriptivas, incluyendo medida de tendencia central y medidas de dispersión. La demanda de los distribuidores presenta una media de 4,446.08 MW y una mediana de 4,271.13 MW, lo que indica una ligera asimetría en la distribución. La varianza es de 1,098,320.90 y la desviación estándar de 1,048.00 MW, reflejando una considerable dispersión de los datos. El valor mínimo registrado es de 2,562.34 MW, mientras que el máximo alcanza los 7,303.73 MW. El primer cuartil (Q1) se encuentra en 3,608.27 MW y el tercer cuartil (Q3) en 5,330.91 MW, lo que muestra que el 50% central de los datos oscila dentro de este rango.

Figura 17

Histograma y curva de densidad ajustada de la potencia de los distribuidores eléctricos del SEIN, tras la aplicación de filtros de depuración.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Tabla 3

Métricas descriptivas de los distribuidores

Medida	Valor	Unidad	Nota
Media	4 446.08	MW	Promedio
Mediana	4 271.13	MW	Valor del medio
Varianza	1 098 320.90	MW^2	M. Dispersión
Desviación estándar	1 048.00	MW	M. Dispersión
Mínimo	2 562.34	MW	Registrado durante la Pandemia
Máximo	7 303.73	MW	Registrado el último año
Primer cuartil (Q1)	3 608.27	MW	-
Tercer cuartil (Q3)	5 330.91	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Tras un análisis detallado de la calidad de los datos, se concluyó que las inconsistencias observadas se deben a errores en el registro, causados por escalas de medición inadecuadas o fallos en la transmisión de información. A pesar de aplicar métodos de reemplazo de outliers como media móvil y mediana local, no se logró una base de datos

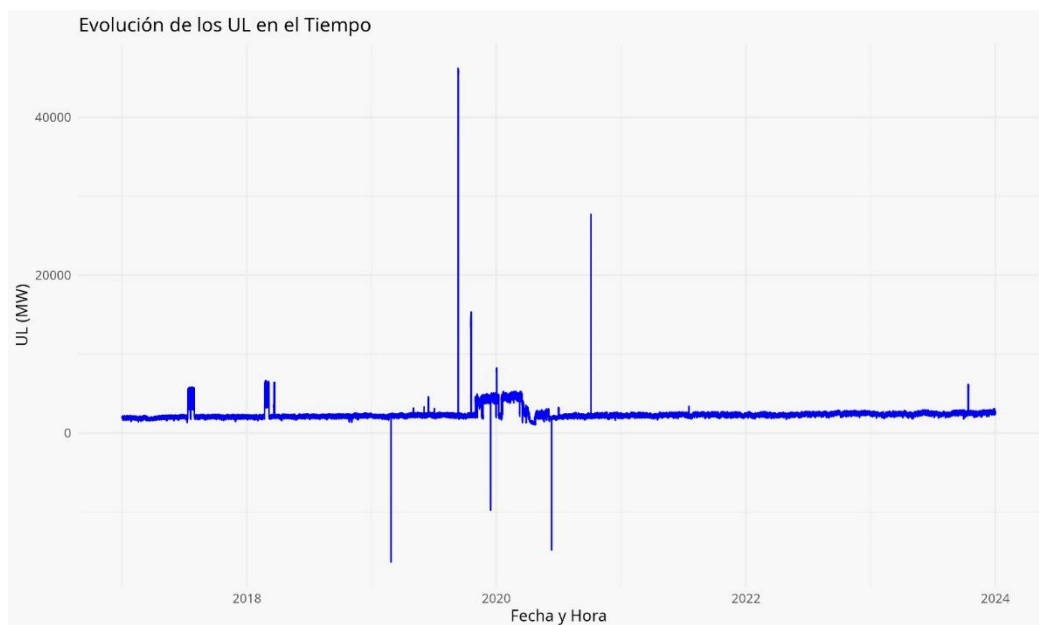
suficientemente fiable para representar adecuadamente la demanda real de los distribuidores.

Usuarios Libres

El análisis de la demanda de los Usuarios Libres presentó un problema similar al de los distribuidores. En la Figura 18, se observa una serie temporal con picos extremos positivos y negativos, lo que introduce distorsiones significativas. El histograma de la Figura 19 evidencia la curva de densidad de color rojo, con una asimetría a la derecha, con alta concentración de datos en valores bajos y una caída abrupta en la densidad hacia valores elevados, indicando la presencia de outliers severos.

Figura 18

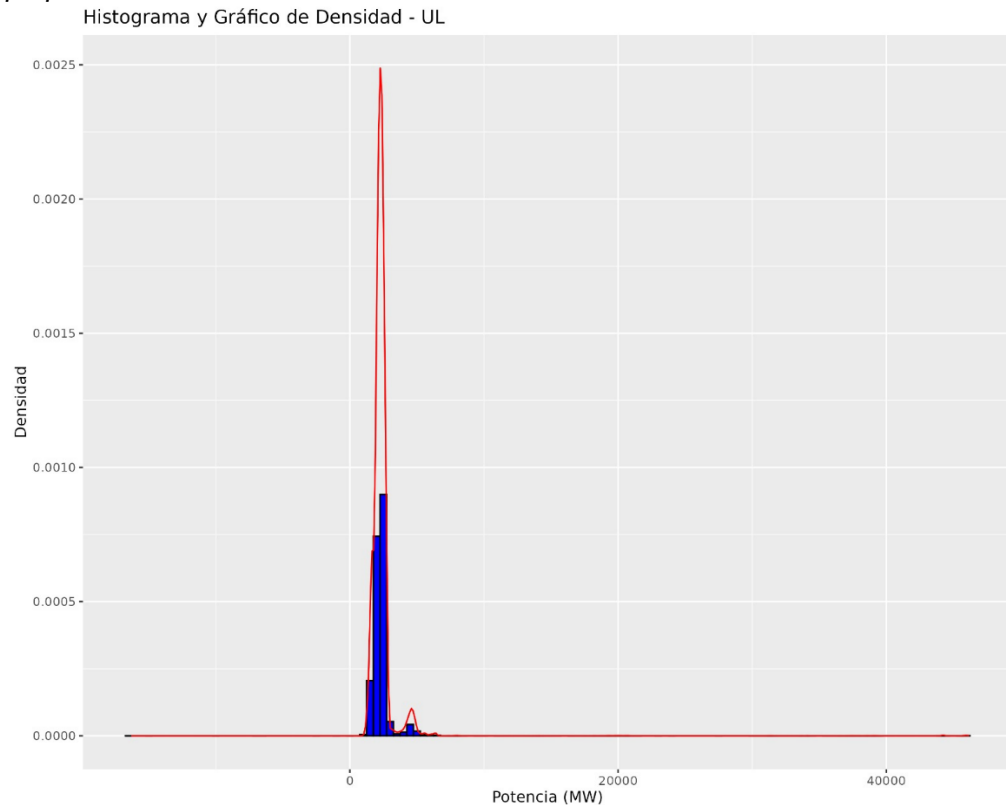
Evolución temporal de la potencia registrada por los usuarios libres del SEIN entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 19

Histograma y curva de densidad de la potencia registrada por los usuarios libres del SEIN luego del preprocesamiento de datos.

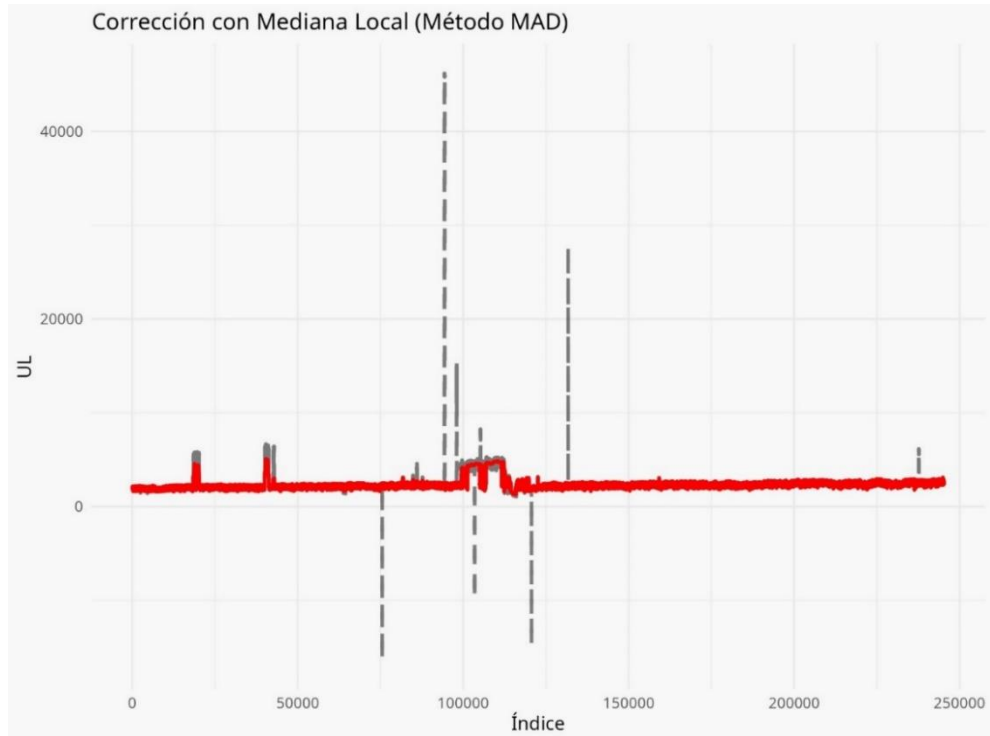


Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Por esta razón se realiza un proceso de limpieza similar al de la distribución, la cual constó únicamente de una etapa, cuya mejor combinación fue la detección de outliers por MAD y su reemplazo por la mediana local. Luego de esto se persistió en el filtrado, sin embargo, no se obtuvieron resultados positivos, se empezó a perder data representativa sin lograr eliminar outliers adecuadamente. La figura 20 muestra la corrección de la data de color rojo y la data filtrada de color gris sombreado. En la figura 21 el resultado final de la data limpia en color azul.

Figura 20

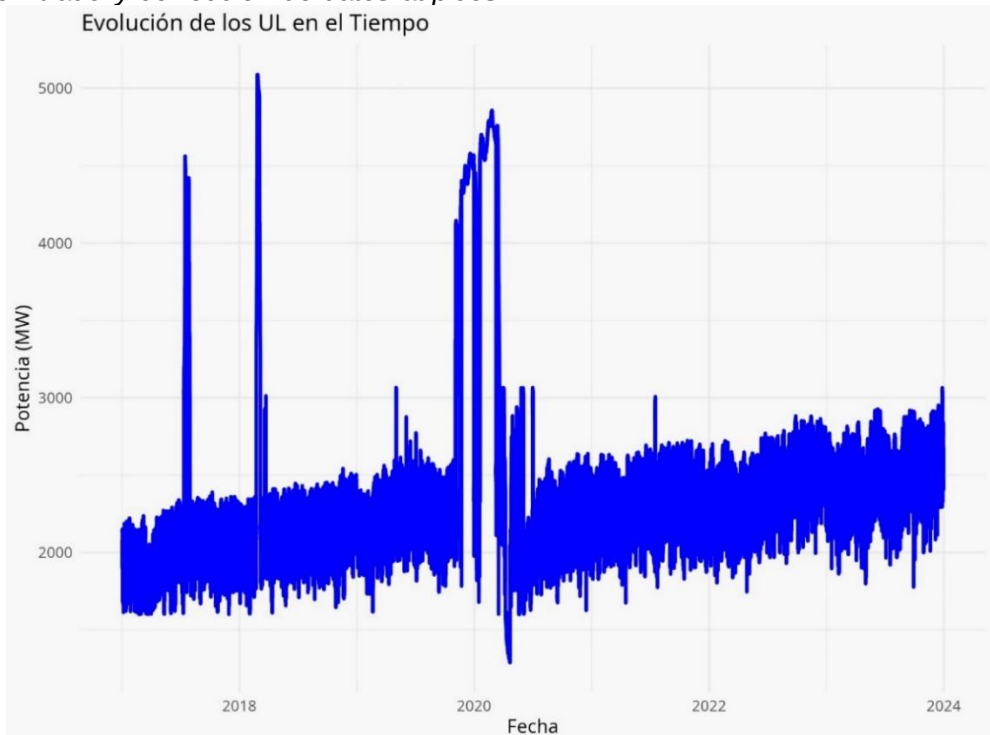
Aplicación del método de filtrado mediante mediana local (MAD) sobre los datos de potencia de usuarios libres del SEIN.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 21

Potencia histórica registrada por los usuarios libres del SEIN entre 2017 y 2024, luego del proceso de filtrado y corrección de datos atípicos.

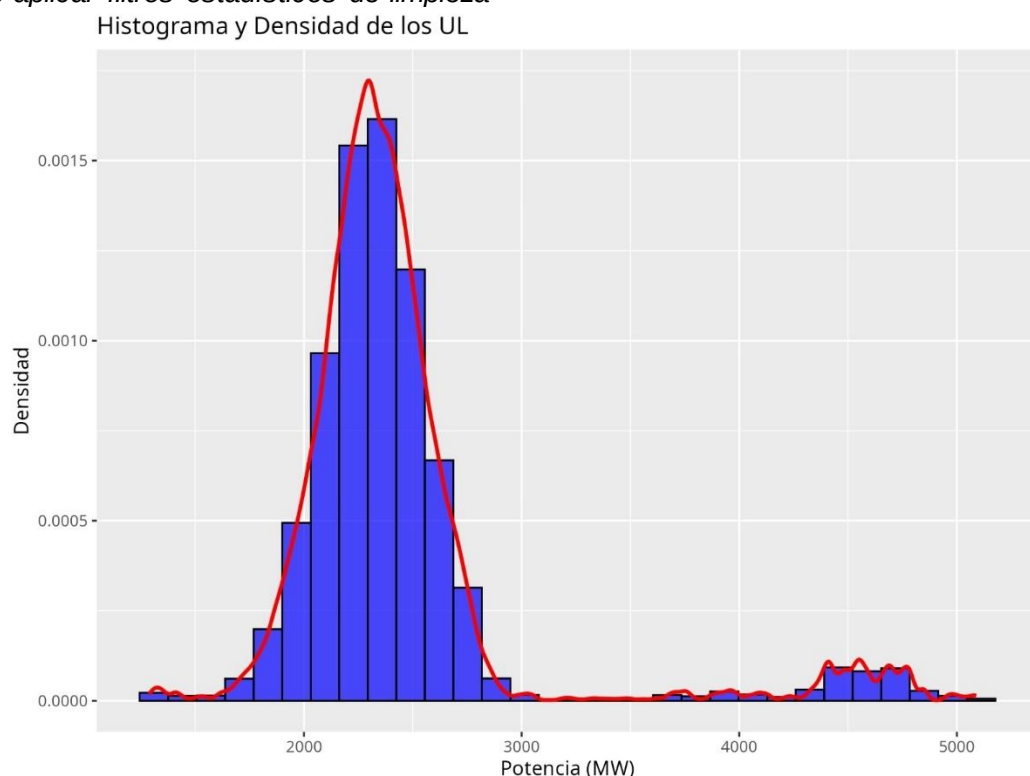


Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Se procede a volver a generar el histograma con esta nueva data limpia, en la figura 22 se visualiza la gráfica con la distribución de la demanda de los usuarios libres (UL) en términos de potencia (MW). Se observa una distribución asimétrica con un pico principal alrededor de los 2,300 MW y una cola hacia valores más altos, lo que sugiere una distribución que aún se mantiene sesgada a la derecha. Además, hay una pequeña concentración de valores en torno a los 4,500 MW, aunque con menor frecuencia. La curva de densidad (en rojo) resalta esta tendencia, indicando que la mayoría de las observaciones se agrupan en el rango de 1,500 a 3,000 MW.

Figura 22

Histograma y curva de densidad de la potencia eléctrica de los usuarios libres del SEIN luego de aplicar filtros estadísticos de limpieza



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La demanda de los usuarios libres de la tabla 4, presenta una media de 2,434.87 MW y una mediana de 2,331.84 MW, lo que indica la ligera asimetría positiva en la distribución. La varianza es de 307,356.66 y la desviación estándar de 554.40 MW, lo que sugiere una menor dispersión en comparación con los distribuidores. El valor mínimo

registrado es de 1,288.54 MW y el máximo alcanza los 5,089.12 MW. El primer cuartil (Q1) se encuentra en 2,176.47 MW y el tercer cuartil (Q3) en 2,504.15 MW, reflejando que el 50% central de los datos se encuentra en un rango más estrecho en comparación con la demanda de los distribuidores.

Tabla 4
Métricas descriptivas de los Usuarios Libres

Medida	Valor	Unidad	Nota
Media	2 434.87	MW	Promedio
Mediana	2 331.84	MW	Valor del medio
Varianza	307 356.66	MW ²	M. Dispersión
Desviación estándar	554.40	MW	M. Dispersión
Mínimo	1 288.54	MW	Registrado durante la Pandemia
Máximo	5 089.12	MW	Registrado el último año
Primer cuartil (Q1)	2 176.47	MW	-
Tercer cuartil (Q3)	2 504.15	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Informe de Evaluación de Operación Diaria (IEOD)

Es un informe diario que publica el COES desde el 2011, a lo largo de estos años este informe ha ido variando en forma y estructura buscando brindar información de mejor calidad. Se toma el 2017 como el primer año de análisis porque a partir de ese año los informes están mejor estructurados incluyendo nuevas métricas e información relevante que se mantiene hasta la fecha.

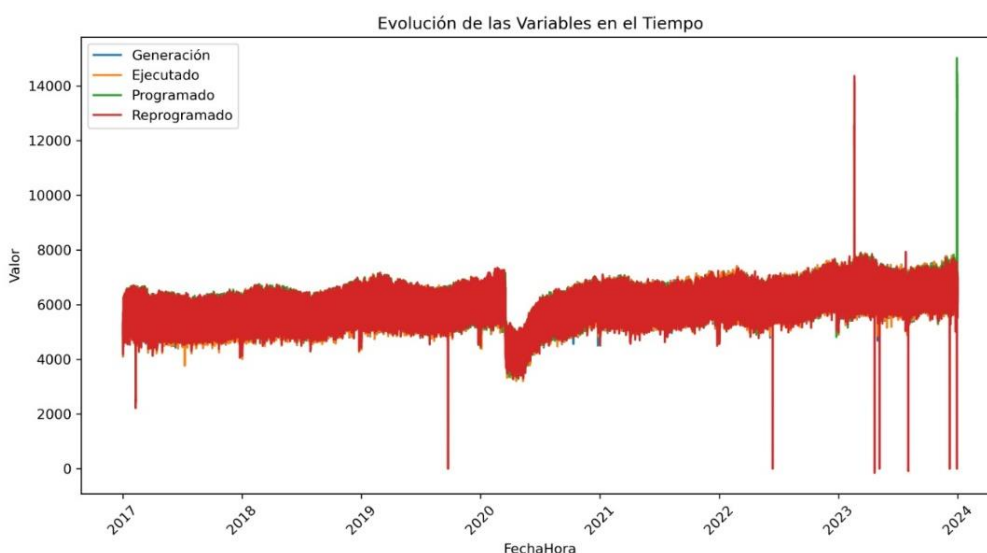
Los informes IEOB están disponibles en diversos formatos (Excel, PDF, .Zip), y contienen datos sobre mantenimientos, operación, despacho, programación diaria y costos marginales. Para este estudio, se trabajó con el Anexo A (anteriormente Anexo 1), que agrupa hojas como Generación por Áreas, Cogeneración, Tipo de Recurso, Costo Marginal, Despacho Ejecutado, Programado, Reprogramado, y Demanda por Áreas, entre otras. Las variables seleccionadas para el análisis son:

- Despacho Ejecutado: Potencia ejecutada y registrada en tiempo real por el COES en intervalos de 30 minutos, posee una leve diferencia con los valores registrados en la generación, pero brinda una idea muy cercana a los que realmente está sucediendo.
- Despacho Programado: Es la potencia programada a ejecutarse según el análisis de consumo de los demandantes y disponibilidad de generación de los ofertantes, que el COES realiza de manera permanente.
- Despacho Reprogramado: Son los cambios que existen durante el transcurso del despacho, por algún cambio, falla, mantenimiento, etc.

En la Figura 23, se observa cómo estas tres variables (Ejecutado: Color naranja, Programado: Color verde, Reprogramado: Color guinda) evolucionan de manera prácticamente idéntica a lo largo del tiempo en comparación con la Generación (color azul). No obstante, se observa un problema de outliers con la data del programado y del reprogramado.

Figura 23

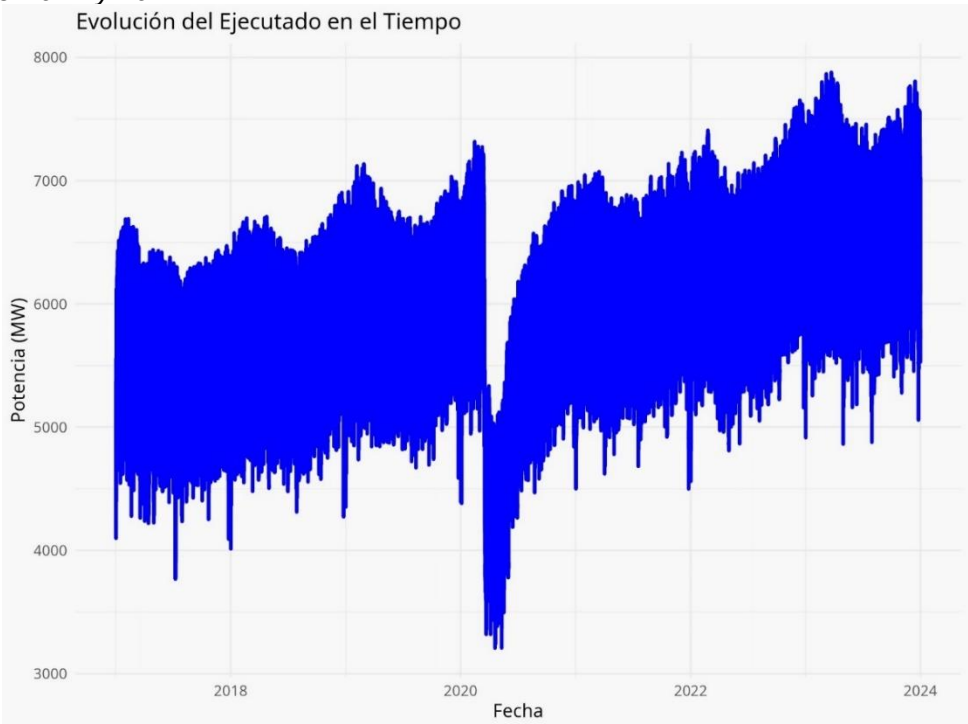
Evolución temporal de la potencia ejecutada, programada, reprogramada y generada en el SEIN entre 2017 y 2024.



Nota: Elaboración propia en Jupyter Notebook con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Se aplicó la misma metodología empleada en variables anteriores para el tratamiento de la Potencia Programada y Potencia Reprogramada. En ambos casos, los valores atípicos se detectaron utilizando la técnica del Z-Score y fueron reemplazados por la media móvil. Las figuras de la 24 a la 29, presentan las series depuradas junto con sus respectivos histogramas y las tablas estadísticas descriptivas correspondientes.

Figura 24
Evolución de la potencia ejecutada en el Sistema Eléctrico Interconectado Nacional (SEIN) entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>.

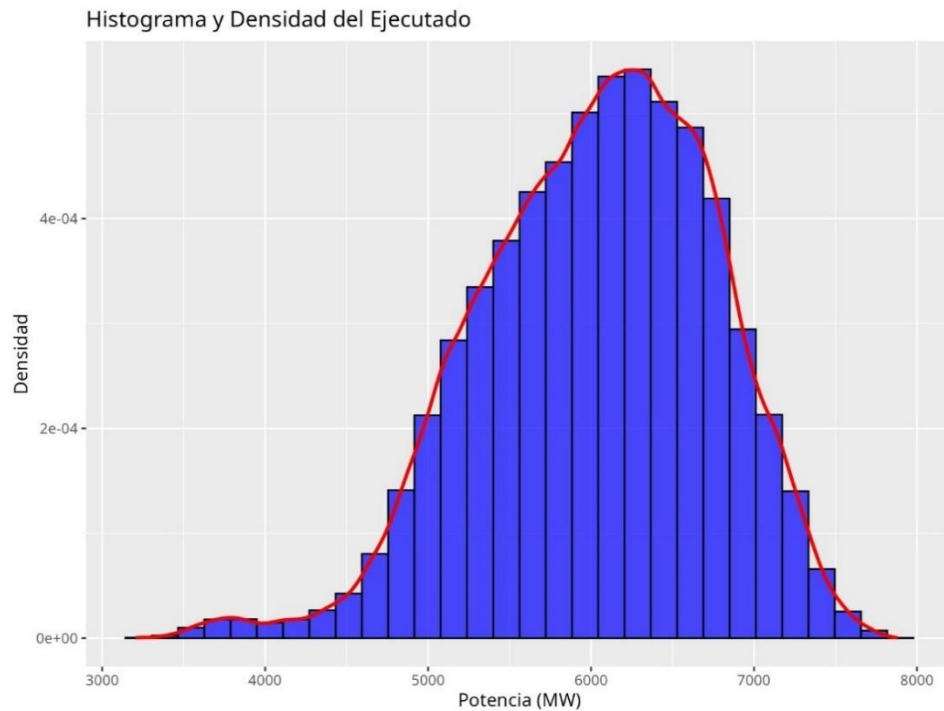
Tabla 5
Métricas descriptivas de la Potencia Ejecutada

Medida	Valor	Unidad	Nota
Media	6 033.70	MW	Promedio
Mediana	6 087.28	MW	Valor del medio
Varianza	504 425.95	MW^2	M. Dispersión
Desviación estándar	710.23	MW	M. Dispersión
Mínimo	3 206.14	MW	Registrado durante la Pandemia
Máximo	7 880.46	MW	Registrado el último año
Primer cuartil (Q1)	5 547.85	MW	-
Tercer cuartil (Q3)	6 561.49	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 25

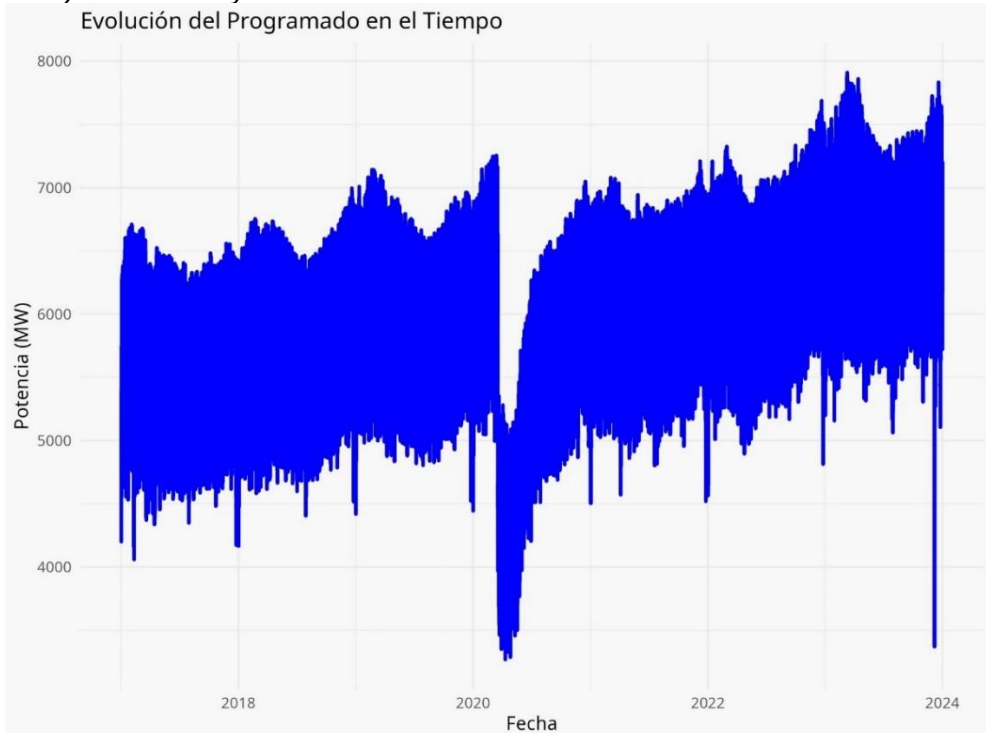
Histograma y curva de densidad de la potencia ejecutada registrada en el SEIN durante el periodo 2017–2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 26

Evolución temporal de la potencia programada en el Sistema Eléctrico Interconectado Nacional (SEIN) entre 2017 y 2024.



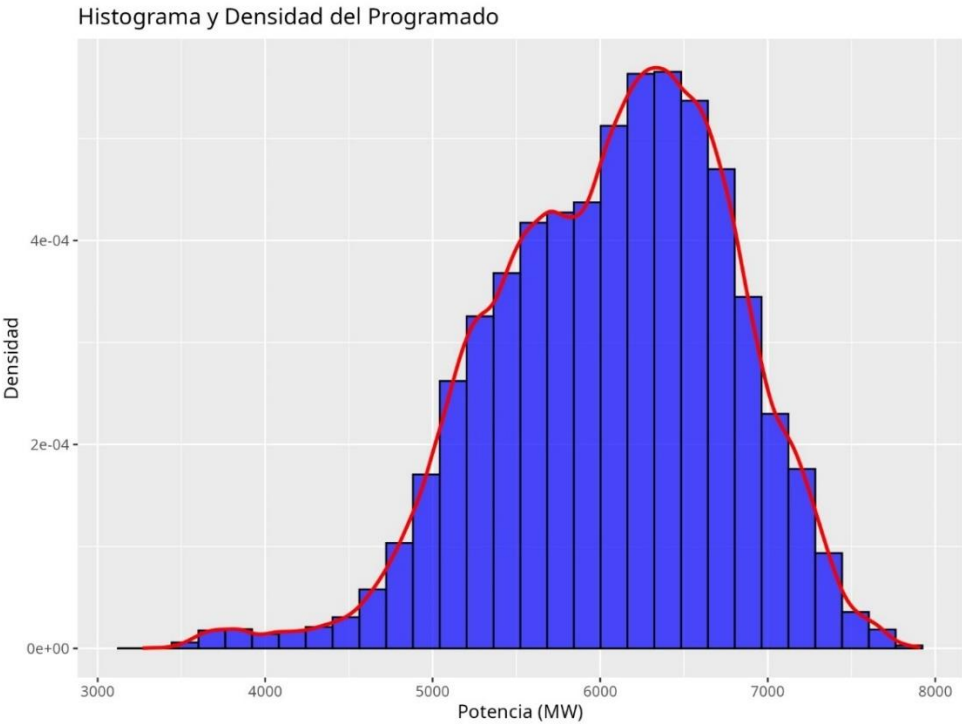
Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Tabla 6
Métricas descriptivas de la Potencia Programada

Medida	Valor	Unidad	Nota
Media	6 069.85	MW	Promedio
Mediana	6 137.60	MW	Valor del medio
Varianza	492 988.80	MW ²	M. Dispersión
Desviación estándar	702.13	MW	M. Dispersión
Mínimo	3 268.02	MW	Registrado durante la Pandemia
Máximo	7 910.31	MW	Registrado el último año
Primer cuartil (Q1)	5 579.89	MW	-
Tercer cuartil (Q3)	6 585.82	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

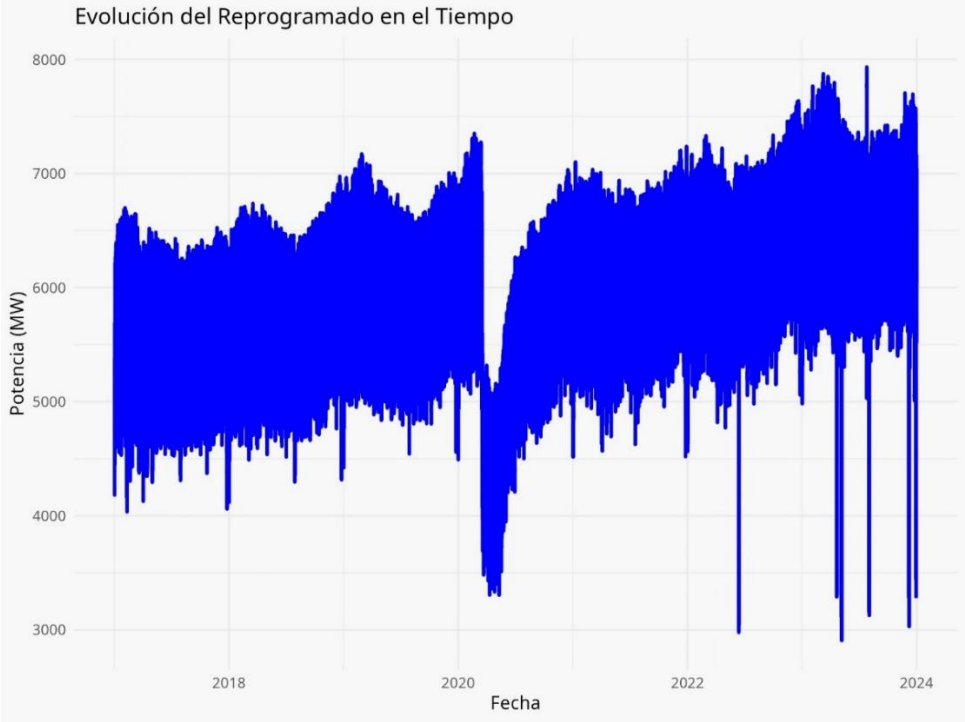
Figura 27
Histograma y curva de densidad de la potencia programada en el SEIN entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 28

Evolución temporal de la potencia reprogramada en el SEIN entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Tabla 7

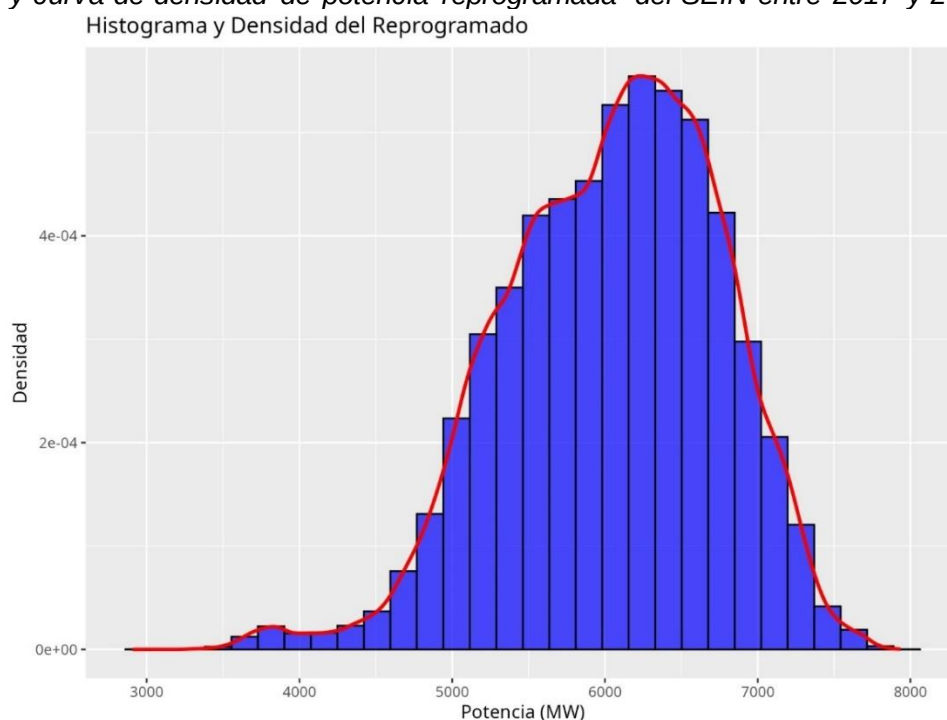
Métricas descriptivas de la Potencia Reprogramada

Medida	Valor	Unidad	Nota
Media	6 043.47	MW	Promedio
Mediana	6 103.59	MW	Valor del medio
Varianza	488 728.02	MW^2	M. Dispersión
Desviación estándar	699.09	MW	M. Dispersión
Mínimo	2 906.59	MW	Registrado durante la Pandemia
Máximo	7 935.11	MW	Registrado el último año
Primer cuartil (Q1)	5 557.11	MW	-
Tercer cuartil (Q3)	6 562.14	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 29

Histograma y curva de densidad de potencia reprogramada del SEIN entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

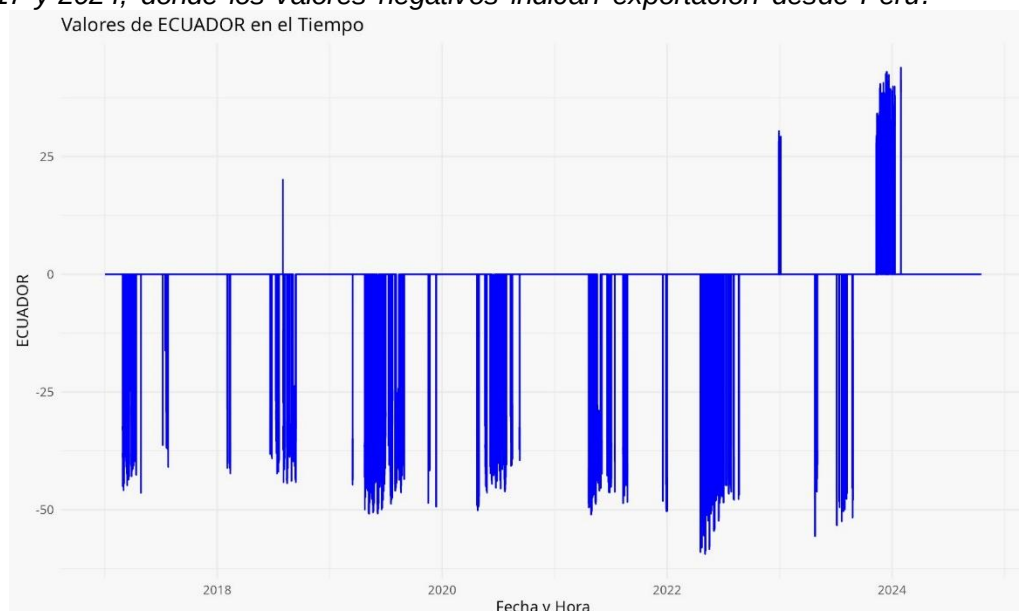
La limpieza de datos permitió mejorar su calidad eliminando valores atípicos. Como resultado, las tres variables analizadas (ejecutada, programada y reprogramada) presentan estadísticas descriptivas similares entre sí y con la generación de potencia, destacando una distribución cercana a la normalidad. No obstante, persiste una ligera asimetría negativa causada por los valores atípicos registrados durante la pandemia del año 2020.

Registros de Contadores de Energía (Intercambios Internacionales)

Los registros de contadores de energía incluyen datos sobre los intercambios internacionales de potencia entre Perú y Ecuador, los cuales afectan la MDC. En la Figura 30, se observa que estos intercambios presentan un comportamiento errático y discontinuo, con picos irregulares en determinados períodos. Esta variabilidad se debe a la naturaleza de los acuerdos de importación y exportación de energía, los cuales responden a factores externos como el estado de las redes interconectadas, la oferta y demanda en ambos países, y la capacidad de transmisión disponible.

Figura 30

Comportamiento de la potencia de intercambio internacional con Ecuador, registrada entre 2017 y 2024, donde los valores negativos indican exportación desde Perú.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

El análisis de esta variable reveló que, aunque los intercambios pueden incidir en la ocurrencia de la MDC, su comportamiento errático y no sistemático dificulta su inclusión en modelos de predicción basados en patrones históricos. El histograma en la figura 31 muestra la curva de densidad de color rojo con una fuerte concentración en torno a cero, lo que indica que la mayoría de los intercambios son nulos. Los valores extremos observados, aunque poco frecuentes, corresponden a eventos significativos, especialmente hacia importaciones.

Tabla 8

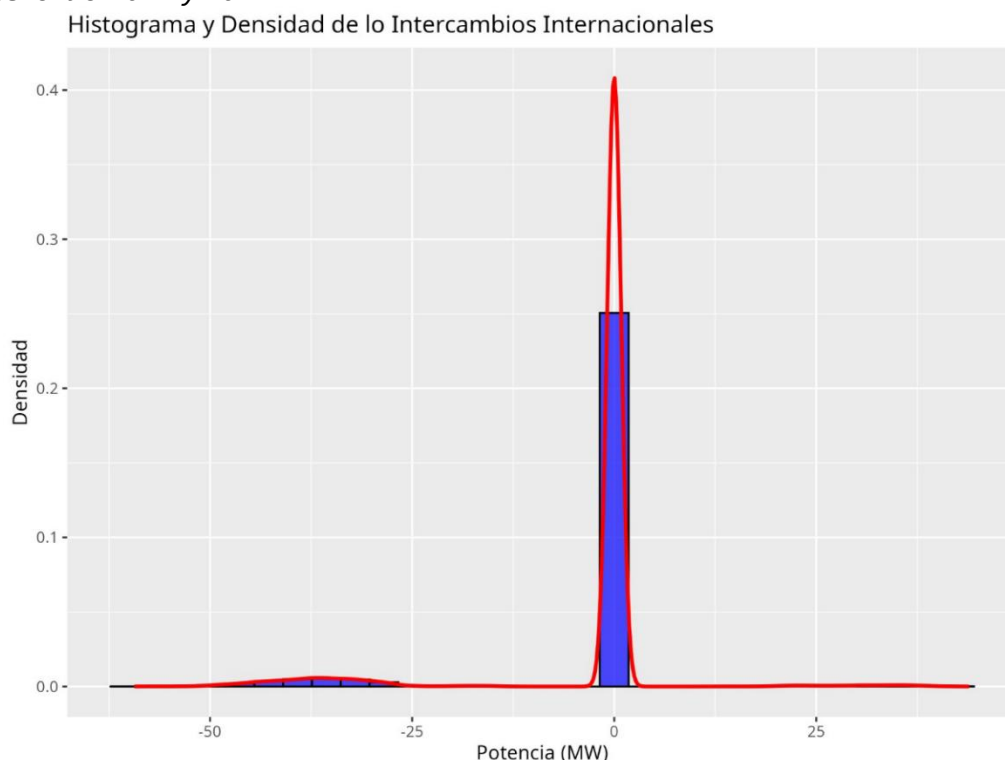
Métricas descriptivas de la Potencia de Intercambio Internacional

Medida	Valor	Unidad	Nota
Media	-2.99	MW	Promedio
Mediana	0	MW	Valor del medio
Varianza	133.85	MW^2	M. Dispersión
Desviación estándar	11.57	MW	M. Dispersión
Mínimo	-59.38	MW	Máxima importación registrada
Máximo	43.91	MW	Máxima exportación registrada
Primer cuartil (Q1)	0	MW	-
Tercer cuartil (Q3)	0	MW	-

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 31

Histograma y curva de densidad de la potencia de intercambio internacional con Ecuador, registrados entre 2017 y 2024.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

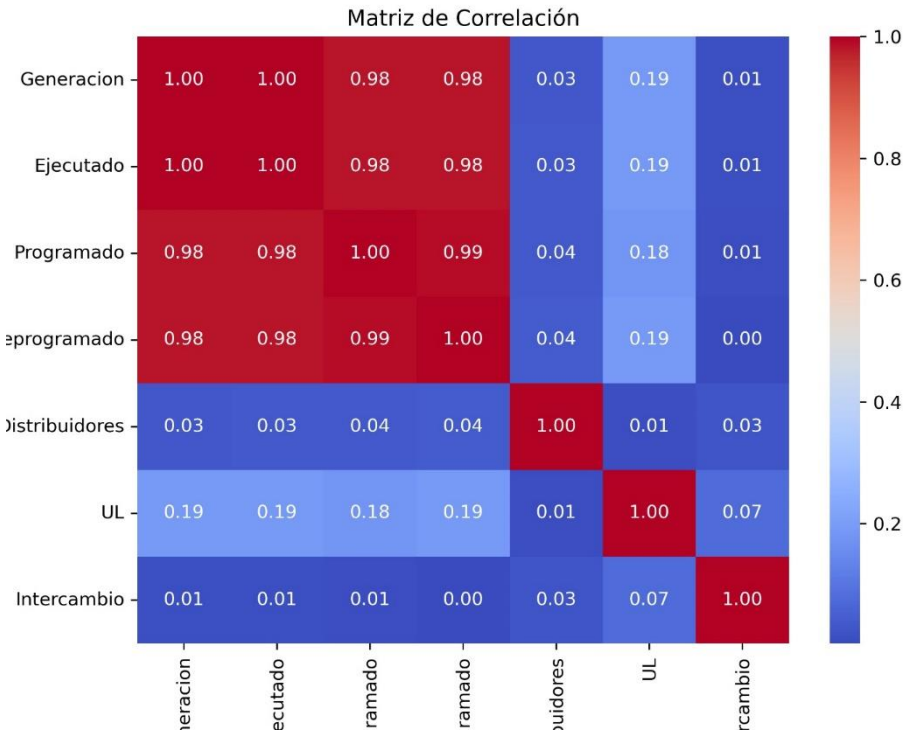
En la tabla 8 se muestra una media de -2.99 MW, cercana a cero, lo que sugiere un leve predominio de importaciones, pero en general un equilibrio entre exportaciones e importaciones. La mediana, así como los cuartiles (Q1 y Q3), tienen valores de 0, indicando que gran parte de los intercambios son nulos, es decir, durante la mayor parte del tiempo no hay intercambio activo. La varianza de 133.85 y la desviación estándar de 11.57 MW reflejan una baja dispersión general, aunque con episodios significativos. La diferencia entre el mínimo (-59.38 MW) y el máximo (43.91 MW) evidencia la presencia ocasional de eventos extremos en ambos sentidos.

3.1.5.3 Análisis de Correlación de Variables

Con todas las variables necesarias previamente depuradas, se procede a realizar un análisis de correlación para evaluar el grado de asociación entre ellas y su relación con la variable objetivo, la potencia de generación (target). La matriz de correlación, presentada en la figura 32, permite identificar las variables más relevantes para el modelo de predicción

de la Máxima Demanda Coincidente (MDC). Para garantizar la coherencia en el análisis, todas las variables han sido estandarizadas con una granularidad de 15 minutos y un rango temporal homogéneo, abarcando desde enero de 2017 hasta diciembre de 2024.

Figura 32
Matriz de correlación de variables energéticas del SEIN entre 2017 y 2024, utilizando coeficientes de correlación de Pearson.



Nota: Elaboración propia en Jupyter Notebook con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Alta correlación de la Generación, Ejecutado, Programado y Reprogramado (≥ 0.98)

- Estas variables muestran una relación prácticamente lineal entre sí, lo que indica que contienen información muy similar.
- Dado que la redundancia entre variables genera problemas en el modelado, se analiza la opción de eliminar algunas de ellas para reducir la complejidad innecesaria del modelo.

Baja correlación de Distribuidores y UL con las demás variables (≤ 0.20 , excepto consigo misma)

- La variable Distribuidores no tiene una relación significativa con Generación ni con sus variables relacionadas (Ejecutado, Programado, Reprogramado, UL).
- A diferencia de Distribuidores, la variable UL presenta una correlación baja pero no totalmente despreciable con la Generación.
- Estas variables no aportan información relevante, lo que vuelve al modelo más complejo sin mejorarlo significativamente.

Intercambio Internacional con las demás variables (< 0.07)

- La baja correlación de Intercambio sugiere que su comportamiento es independiente de la generación y planificación de demanda.
- Ya que no tiene un significado físico relevante en el sistema eléctrico que justifique su uso, su inclusión en el modelo no es necesaria.

3.1.5.4 Herramientas Utilizadas para el EDA

Para la ejecución del Análisis Exploratorio de Datos (EDA) se utilizaron herramientas especializadas en el procesamiento, manipulación y visualización de datos, con el objetivo de evaluar la calidad, la distribución y la relación entre las variables analizadas. El análisis se realizó utilizando Python y R, aprovechando sus respectivas ventajas en el manejo y visualización de datos.

En términos de entornos de desarrollo, se trabajó en Jupyter Notebook bajo un entorno Conda con un kernel específico configurado para este estudio, lo que permitió integrar diferentes librerías y gestionar dependencias de manera eficiente. Para los análisis en R, se utilizó RStudio, aprovechando su capacidad para la manipulación de datos con librerías especializadas.

En Python, se emplearon las siguientes librerías:

- Pandas: Manipulación de datos estructurados.
- NumPy: Operaciones numéricas y estadística básica.
- Matplotlib y Seaborn: Visualización de datos mediante gráficos de series temporales e histogramas.
- Statsmodels: Aplicación de modelos estadísticos básicos en el EDA.

En R, las librerías utilizadas fueron:

- dplyr: Transformación y manipulación de datos.
- ggplot2: Visualización avanzada de datos.
- zoo: Manejo de series temporales.
- readxl y readr: Lectura y procesamiento de archivos en Excel y CSV.

Los datos del COES fueron transformados a CSV para facilitar su manipulación. Posteriormente, se utilizaron principalmente gráficos de series temporales e histogramas para analizar tendencias, estacionalidad y valores atípicos:

- Gráficos de series temporales, que permitieron identificar tendencias y patrones estacionales en la generación de potencia.
- Histogramas, utilizados para analizar la distribución de las variables y detectar posibles asimetrías o valores atípicos.

3.1.5.5 Conclusiones del EDA

El análisis de la generación de potencia del SEIN confirmó una tendencia creciente a lo largo de los años, coherente con el aumento de la demanda energética asociado al desarrollo tecnológico y económico del país. También se identificó una estacionalidad semanal, reflejo de ciclos operacionales en el sistema eléctrico. Aunque existen patrones definidos, se observaron variaciones significativas dentro de un mismo día, con oscilaciones entre valores máximos y mínimos en cortos intervalos.

3.1.6 Decisiones Tomadas en Base al EDA

- La demanda de Distribuidores y Usuarios Libres (UL) son descartadas por deficiencias en la calidad de los datos y baja correlación con la generación de potencia. Los registros presentaban valores atípicos que no pudieron corregirse mediante preprocesamiento, limitando su utilidad en el modelo.
- El despacho programado y el despacho reprogramado también son descartadas por alta correlación redundante. Estas variables presentaban una alta correlación con la generación de potencia y entre sí, lo que generaba redundancia. Para evitar modelos innecesariamente complejos, se seleccionó solo la Potencia Ejecutada, ya que es la variable más representativa del comportamiento real del sistema eléctrico.
- Intercambios Internacionales: Descartados por su comportamiento errático. Aunque estos eventos pueden afectar la ocurrencia de la MDC —por ejemplo, adelantándola o desplazándola a otro horario—, su imprevisibilidad impide incorporarlos en un modelo basado en patrones consistentes.
- La predicción se realizará con una frecuencia semanal, dado que el patrón encontrado se mantiene estable a lo largo del tiempo, independientemente de la tendencia general. Este intervalo resulta más eficiente que una predicción diaria, que requeriría técnicas complejas con alta frecuencia, y más ágil que una predicción mensual, que podría reaccionar tarde ante eventos atípicos y comprometer la precisión del modelo.
- Para el modelado se eligieron dos variables: la potencia de Generación y la potencia Ejecutada. La primera se publica mensualmente y refleja el valor real del target, mientras que la segunda se actualiza diariamente, lo cual permite enriquecer el modelo semanalmente sin depender exclusivamente del dato mensual. Esto facilita una respuesta más rápida ante eventos anómalos.

3.2 Desarrollo de Modelos de Predicción

Se busca predecir la generación de potencia eléctrica del SEIN, con una granularidad de 15 minutos. Para esto se plantean 3 métodos diferentes de modelado:

- SARIMA.
- Prophet.
- Long short-term memory (LSTM).

El objetivo es evaluar cuál de los tres métodos propuestos predice con mayor precisión la generación de potencia eléctrica del SEIN, utilizando únicamente el histórico de dicha variable. Por simplicidad y para garantizar fiabilidad, se selecciona la generación de potencia como única variable, cumpliendo el rol tanto de entrada (feature) como de salida (target). Esto permite centrarse en el rendimiento de los modelos sin introducir complejidad innecesaria. El periodo de análisis abarca desde enero de 2017 hasta junio de 2024, con el fin de asegurar una representación adecuada de la infraestructura eléctrica actual, conforme a lo detallado en capítulos anteriores.

3.2.1 SARIMA

El modelo SARIMA se implementó en R usando RStudio, aprovechando sus fortalezas estadísticas. Se importaron las librerías forecast, ggplot2 y readr, y se cargó la data de generación de potencia en formato CSV. Con una granularidad de 15 minutos, se generó una serie temporal con frecuencia diaria (96 observaciones por día). La partición de los datos se realizó en dos conjuntos: entrenamiento (enero 2017 – diciembre 2022) y validación (enero 2023 – junio 2024). El modelo se ajustó automáticamente al conjunto de entrenamiento y se guardó en formato .RData o .rds. Finalmente, se analizaron los residuos para verificar la presencia de ruido blanco. El código completo se encuentra en el Anexo B y los resultados del modelo se presentan en la figura 33, donde se detallan sus componentes y métricas de error sobre el conjunto de entrenamiento.

Figura 33

Resultados del modelo SARIMA aplicado a la serie de potencia eléctrica, con visualización de coeficientes, AIC/BIC y métricas de error.

```
> modelo_sarima <- auto.arima(serie_ts, seasonal = TRUE)
> summary(modelo_sarima)
Series: serie_ts
ARIMA(4,0,2)(0,1,0)[96] with drift

Coefficients:
      ar1      ar2      ar3      ar4      ma1      ma2      drift
 1.6610 -0.5691 -0.0426 -0.0538 -1.0147  0.2172  0.0057
s.e.  0.0098  0.0157  0.0035  0.0036  0.0098  0.0093  0.0283

sigma^2 = 3110: log likelihood = -4672597
AIC=9345210 AICc=9345210 BIC=9345303

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -5.438299e-05 55.76455 33.47497 -0.008564863 0.830357 0.186489 0.0002343811
> |
```

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

El modelo ajustado tiene los siguientes componentes:

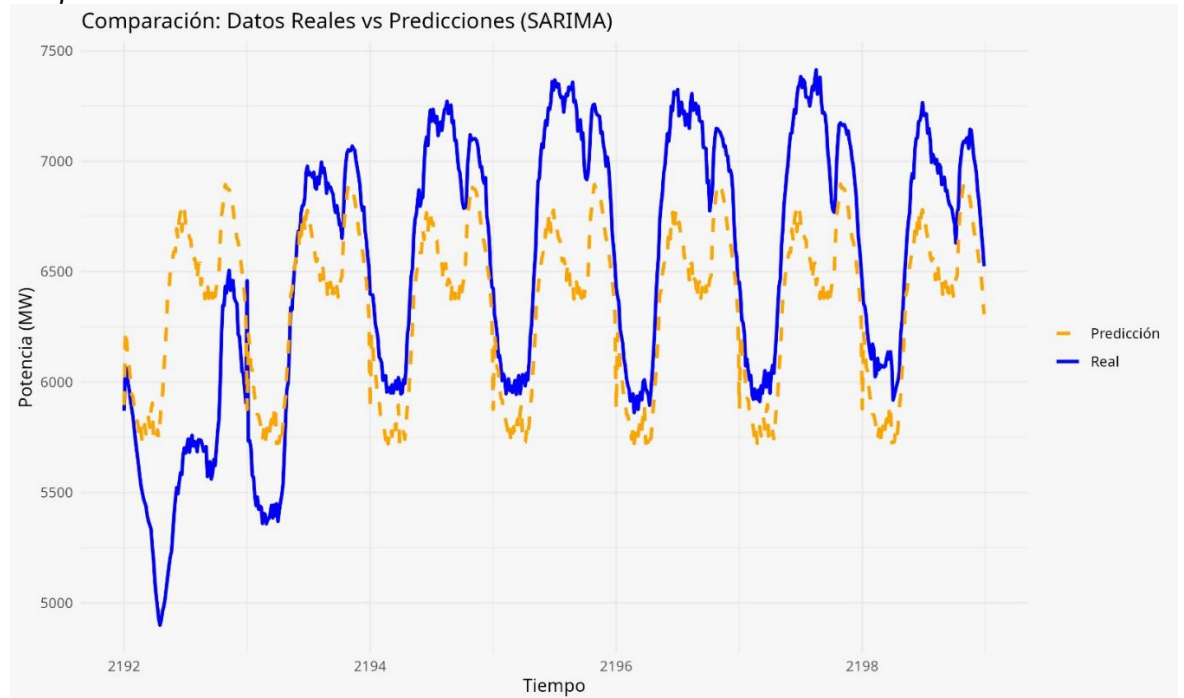
- ARIMA (4,0,2): El 4 (AR) indica que el modelo usa 4 términos autorregresivos (relación lineal con los valores pasados). El 0 (D) indica que no es necesario diferenciar la serie temporal para hacerla estacionaria. Y el 2 (MA), usa 2 términos de media móvil para modelar el ruido blanco.
- (0,1,0): Describe la parte estacional del modelo. El 0 (AR estacional), no usa términos autorregresivos estacionales. El 1 (D estacional) diferenciación estacional de grado 1 para eliminar la estacionalidad. Y el 0 (MA estacional) no usa términos de media móvil estacional.
- Drift: Este término indica que el modelo incluye una tendencia lineal para capturar cambios globales en la serie. *drift* = 0.0057 representa un aumento constante en la serie temporal.
- AR (Autorregresivos): *ar1* = 1.6610, *ar2* = -0.5691, *ar3* = -0.0426, *ar4* = -0.0538. Los coeficientes indican cuánto de los valores pasados afecta al valor actual. Los valores positivos indican que los aumentos pasados tienen un efecto positivo, mientras que los negativos sugieren un efecto opuesto.

- MA (Media Móvil): $ma1 = -1.0147$, $ma2 = 0.2172$. Los coeficientes representan cuánto del error pasado afecta al valor actual.
- ME (Mean Error): Promedio del error residual. El error medio residual (ME) fue de $-5.43829e - 05$, valor cercano a cero, lo cual indica ausencia de sesgo sistemático en el modelo.
- RMSE (Root Mean Square Error): Mide el error promedio en las predicciones. Menor es mejor. Aquí: 55.76455.
- MAE (Mean Absolute Error): Error absoluto promedio, menos sensible a valores extremos que RMSE. Aquí: 33.47497.
- MPE (Mean Percentage Error): Error porcentual promedio. Idealmente cercano a 0. Aquí: -0.0085% : Muy bajo.
- MAPE (Mean Absolute Percentage Error): Error porcentual absoluto promedio de 0.830357% , considerado excelente al encontrarse por debajo del 1% .
- MASE (Mean Absolute Scaled Error): Compara el error absoluto promedio del modelo con un modelo base. Aquí: 0.186489: Excelente (valores menores a 1 indican mejor desempeño que un modelo base).
- ACF1: Autocorrelación del residuo en el lag 1. Aquí: 0.0002343811: Muy bajo, lo que indica que los residuos no tienen correlación a corto plazo.

En la figura 34 se observa la predicción del modelo en color amarillo y de color azul la data real para la primera semana del conjunto de prueba. Este intervalo reducido permite evaluar visualmente el comportamiento detallado de la predicción frente a los datos reales.

Figura 34

Comparación semanal entre los datos reales de potencia eléctrica y las predicciones generadas por el modelo SARIMA.

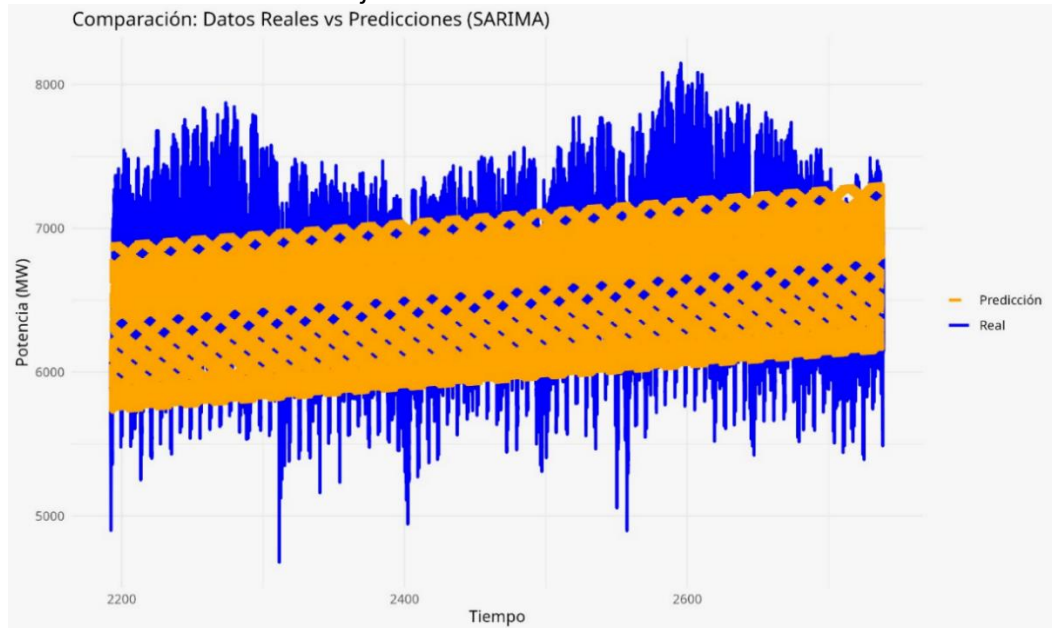


Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La figura 35 muestra la predicción del modelo sobre todo el conjunto de prueba, cubriendo un intervalo más amplio, donde las líneas de color azul es la data real y las líneas de color amarillo es la predicción realizada. Tanto esta figura como la anterior evidencian un bajo desempeño predictivo, lo cual indica que el modelo no logra capturar adecuadamente el comportamiento de la serie para los fines del estudio.

Figura 35

Comparación entre los valores reales de potencia eléctrica y las predicciones generadas por el modelo SARIMA sobre el conjunto de validación.



Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La figura 36 presenta los resultados del test de Ljung-Box, utilizado para verificar si los residuos del modelo se comportan como ruido blanco.

Figura 36

Resultado del test Ljung-Box aplicado a los residuos del modelo SARIMA para evaluar la autocorrelación.

```
> checkresiduals(modelo_sarima)

Ljung-Box test

data: Residuals from ARIMA(4,0,2)(0,1,0)[96] with drift
Q* = 238174, df = 186, p-value < 2.2e-16

Model df: 6. Total lags used: 192
```

Nota: Elaboración propia en RStudio con datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Los resultados del test de Ljung-Box ($Q = 238174$, $p\text{-value} < 2.2e-16$) indican una correlación significativa en los residuos, lo cual descarta que se comporten como ruido blanco. Esto sugiere que el modelo no logra capturar completamente la estructura de la serie temporal, dejando patrones residuales sin modelar. En consecuencia, los residuos contienen información adicional que podría aprovecharse para mejorar la precisión del modelo.

3.2.2 Prophet

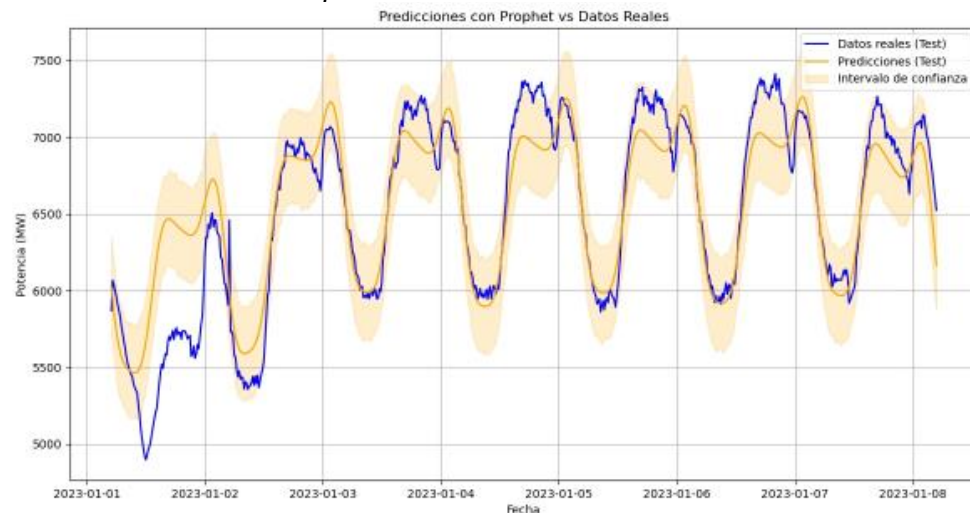
A diferencia de SARIMA, el modelo Prophet se desarrolla en Python utilizando Jupyter Notebook, lo que permite una interacción rápida y robusta. Para su ejecución, se configuró un entorno conda con un kernel exclusivo.

El proceso inicia con la importación de librerías como pandas, prophet, sklearn, matplotlib, numpy, joblib y os. Se carga el mismo archivo CSV con datos de generación cada 15 minutos y se renombran las columnas conforme a los requerimientos de Prophet ('time' como 'ds' y potencia como 'y'). Los datos se dividen en dos conjuntos: entrenamiento ('train', ene. 2017 - dic. 2022) y validación ('validation', ene. 2023 - jun. 2024). El modelo Prophet se inicializa con estacionalidad aditiva (diaria, semanal y anual) y se entrena con la data de entrenamiento. Finalmente, se guarda en formato '.pkl', se generan predicciones sobre el conjunto de prueba y se evalúa el rendimiento del modelo. El código completo se detalla en el Anexo C.

La figura 37 muestra la predicción para la primera semana del conjunto de validación, la línea de color azul es la data real de la generación, la línea de color amarillo es la predicción del modelo y la región naranja sombreada representa el intervalo de confianza de dicha predicción. Esto permite observar con claridad el comportamiento del modelo en un intervalo reducido. Se evidencian errores significativos, con valores que incluso exceden el intervalo de confianza, lo que sugiere una mala generalización. En la figura 38, se observa la predicción sobre todo el conjunto de prueba, donde el intervalo de confianza se amplía progresivamente con el tiempo, reflejando una creciente incertidumbre a medida que avanzan los años.

Figura 37

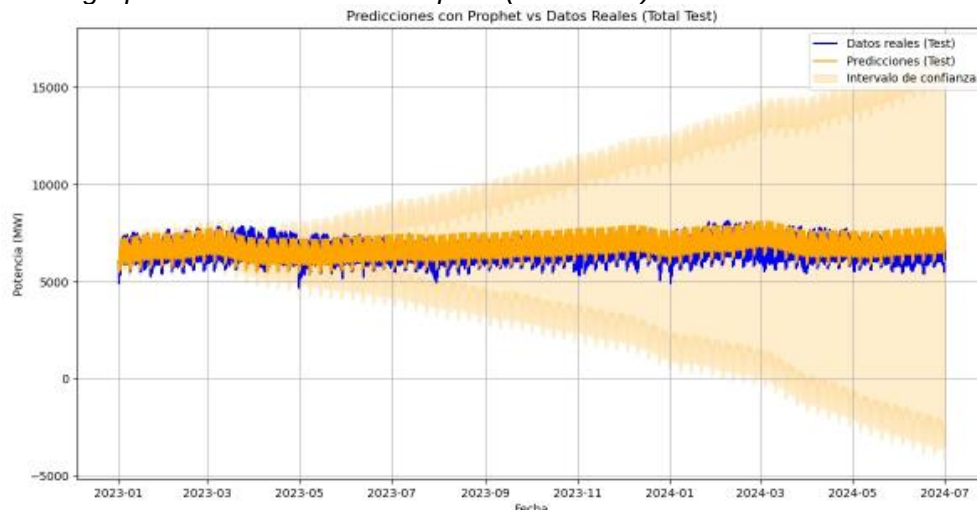
Predicción semanal del modelo Prophet con intervalo de confianza.



Nota: Elaboración propia en Jupyter Notebook con el modelo Prophet de Meta y datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 38

Predicción de largo plazo con el modelo Prophet (validación).



Nota: Elaboración propia en Jupyter Notebook con el modelo Prophet de Meta y datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

A nivel visual, el modelo Prophet logra una mejor generalización que SARIMA; sin embargo, no es lo suficientemente robusto ni confiable. Los picos y valles de la serie real presentan irregularidades que Prophet no logra capturar adecuadamente, generando predicciones suavizadas. Además, se observa que durante los días domingo, los valores reales superan el intervalo de confianza del modelo. Esta incapacidad para representar correctamente los patrones extremos lo convierte en una opción limitada para aplicaciones donde se requiere alta precisión.

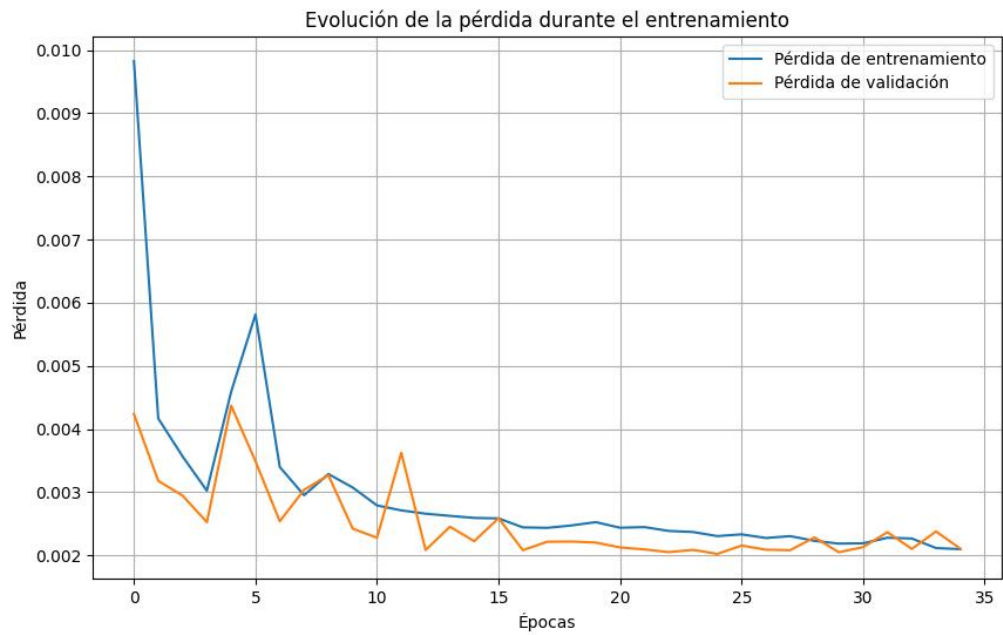
3.2.3 Long short-term memory (LSTM)

De manera similar al modelado en Prophet, se trabaja con Python y Jupyter Notebook, con un entorno conda y kernel exclusivo para esta labor. Se inicia importando las librerías pandas, numpy, sklearn, matplotlib y keras. Se carga el mismo archivo CSV de generación de potencia, se escala la data en un rango de 0 a 1, para mayor facilidad de trabajo con estas bibliotecas. Se genera una función de secuencia para el modelo, con un tamaño de ventana de 672 datos, que representan una semana en la serie temporal, esta secuencia acondiciona los datos para ser usados en el entrenamiento y validación. Se parte la data en testeo y train, donde la validación es el 15% del total de la data y la data de entrenamiento es el 85% de los datos. Se estructura la arquitectura del modelo con dos capas LSTM de 50 neuronas cada una y con un dropout, para evitar el overfitting, del 20%. Se compila el modelo con el optimizador 'adam' y usa el error cuadrático como función de pérdida o también conocido como 'loss'. Luego se entrena con 50 épocas y un tamaño de batch de 32 secuencias, también se usa Early Stopping, que es un callback que permite dejar de entrenar el modelo, y usar el modelo hasta donde tuvo la mejor generalización, si después de ciertas épocas establecidas la función de pérdida de 'test' no logra reducir, evitando extender innecesariamente el entrenamiento y evitando un posible overfitting. Se evalúa el rendimiento y se guarda el modelo en formato '.h5'.

La figura 39 muestra cómo el modelo reduce la pérdida ('loss') tanto en los datos de entrenamiento (línea azul) como en los de validación (línea naranja). Se observa que, después de cierto punto, seguir entrenando no mejora necesariamente la generalización y podría inducir overfitting si no se controla adecuadamente. Posteriormente, se realiza una predicción de una semana, mostrada en la figura 40, donde se comparan los valores predichos (línea naranja) con los reales (línea azul).

Figura 39

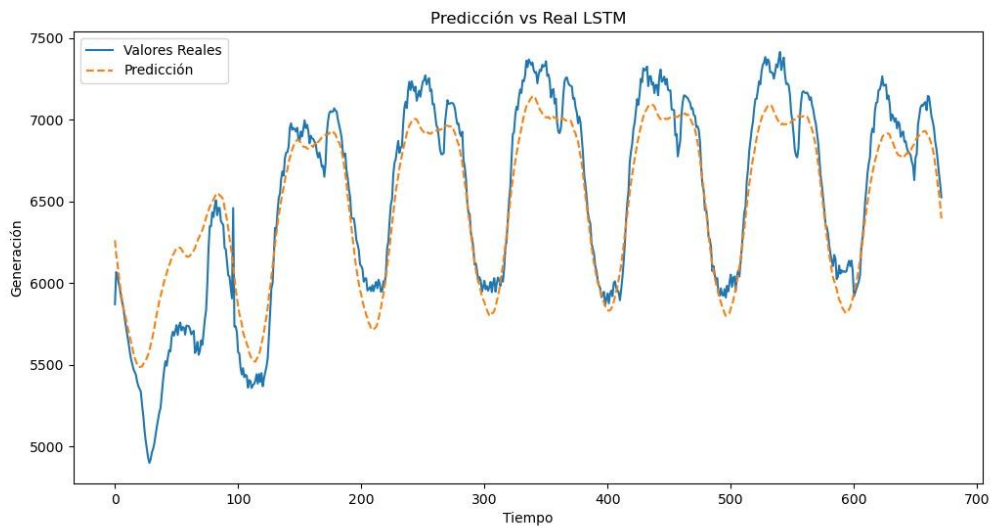
Evolución de la pérdida de entrenamiento y validación del modelo LSTM.



Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos procesados del SEIN (COES, 2024).

Figura 40

Comparación de Predicción y Real (modelo LSTM).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos extraídos del SEIN (COES, 2024).

Los tres modelos fueron entrenados y validados con el mismo conjunto de datos.

Se comparó su desempeño utilizando las métricas MAE , $RMSE$, $MAPE$ y R^2 .

Tabla 9*Performance de los modelos.*

Modelo	MAE	RMSE	MAPE	R Squared
SARIMA	324.2381	409.2637	4.7584	0.4968
Prophet	272.3571	344.9253	4.1633	0.6425
LSTM	180.9068	230.8566	2.8657	0.8610

Nota: MAE = Mean Absolute Error, RMSE = Root Mean Square Error, MAPE = Mean Absolute Percentage Error, R² = Coeficiente de determinación. Estadísticas calculadas con RStudio

La Tabla 9 permite concluir de manera objetiva que **el modelo LSTM es el más robusto y preciso entre las opciones evaluadas**, mostrando un menor error y una mejor capacidad de generalización en relación con la naturaleza de los datos y el caso de estudio abordado. Las diferencias en las métricas de evaluación son lo suficientemente significativas y justificadas como para descartar por completo los otros modelos probados, dado su desempeño inferior en la predicción de la generación de potencia del SEIN.

Seleccionado el modelo LSTM como la mejor alternativa, se evaluaron distintas configuraciones arquitectónicas con el fin de identificar la estructura que optimice su rendimiento predictivo. La Tabla 10 muestra las variantes consideradas, con cambios en el número de capas, neuronas y tasa de dropout. El modelo inicial, definido empíricamente, contó con dos capas de 50 neuronas y una tasa de dropout del 20%. Desde esta base, se probarán configuraciones alternativas para analizar su impacto en la precisión y estabilidad de las predicciones.

Tabla 10*LSTM con diferentes arquitecturas.*

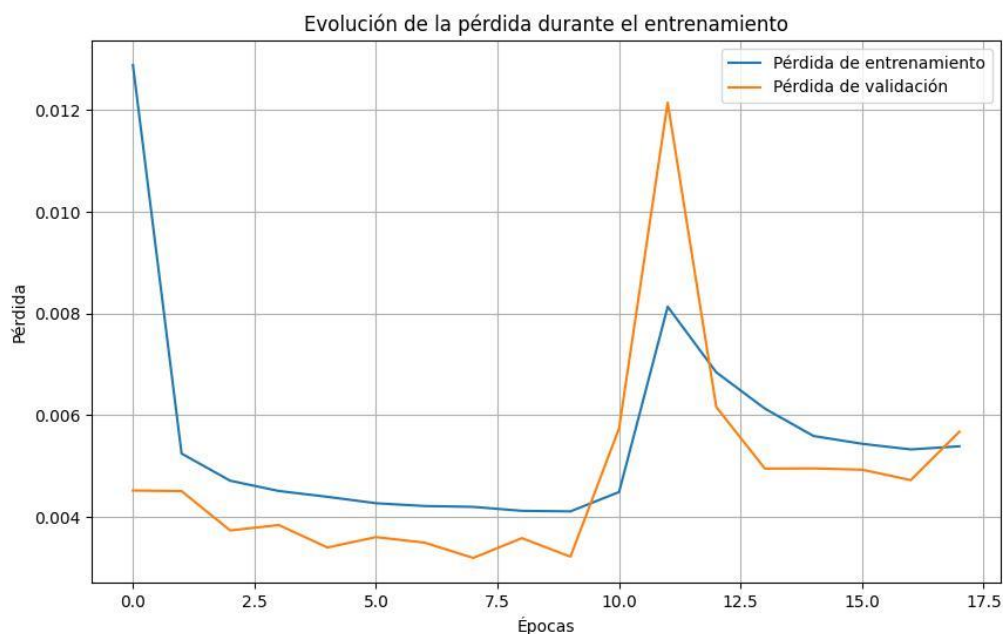
Modelo	Capa 1	Capa 2	Capa 3	Capa 4	Dropout (%)
Modelo A	50	50	0	0	20%
Modelo B	64	32	16	0	20%
Modelo C	64	32	16	8	30%

Nota: La única diferencia entre los modelos es su arquitectura

El entrenamiento de cada uno de estos modelos se llevó a cabo siguiendo el procedimiento previamente descrito. Una vez completado el proceso de entrenamiento, se

generaron visualizaciones que permiten evaluar el desempeño de cada modelo. En particular, se presentan gráficos que ilustran la evolución de la función de pérdida durante el entrenamiento, así como la comparación entre las predicciones generadas y los valores reales en un intervalo representativo de una semana. Estas representaciones gráficas, mostradas en las figuras 41 y 43, muestran la pérdida de la data de entrenamiento con líneas de color azul y la pérdida de la data de validación con la línea de color naranja. Las figuras 42 y 44 muestran la predicción semanal en líneas discontinuas de color naranja y la data real en una línea sólida de color azul; esto permite analizar la estabilidad del modelo, su capacidad de generalización y su desempeño en la predicción de la generación de potencia del SEIN. Se presentan únicamente el modelo B y modelo C, ya que el modelo A, ya fue evaluado en la figura 40 y tabla 9.

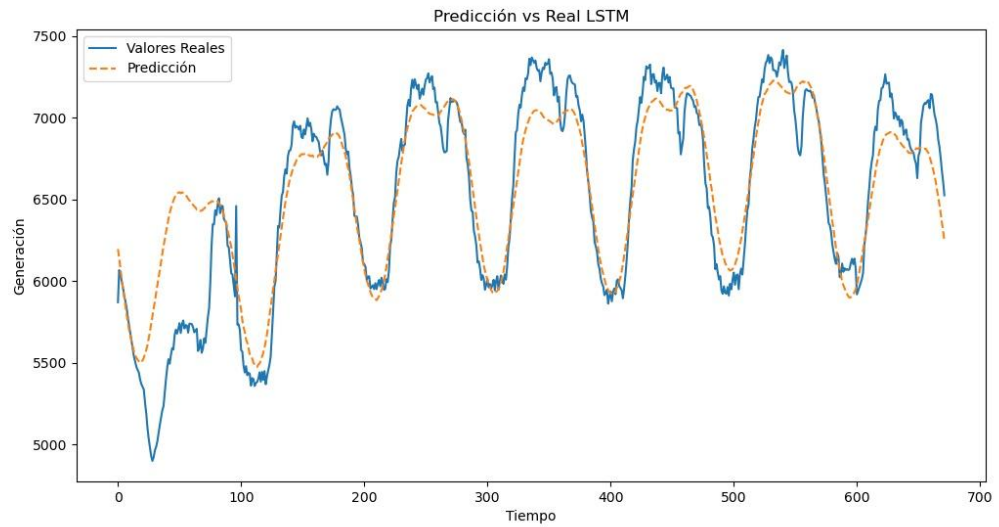
Figura 41
Pérdida de entrenamiento del Modelo B (LSTM).



Nota: Elaboración propia en Jupyter Notebook (Python 3.10)

Figura 42

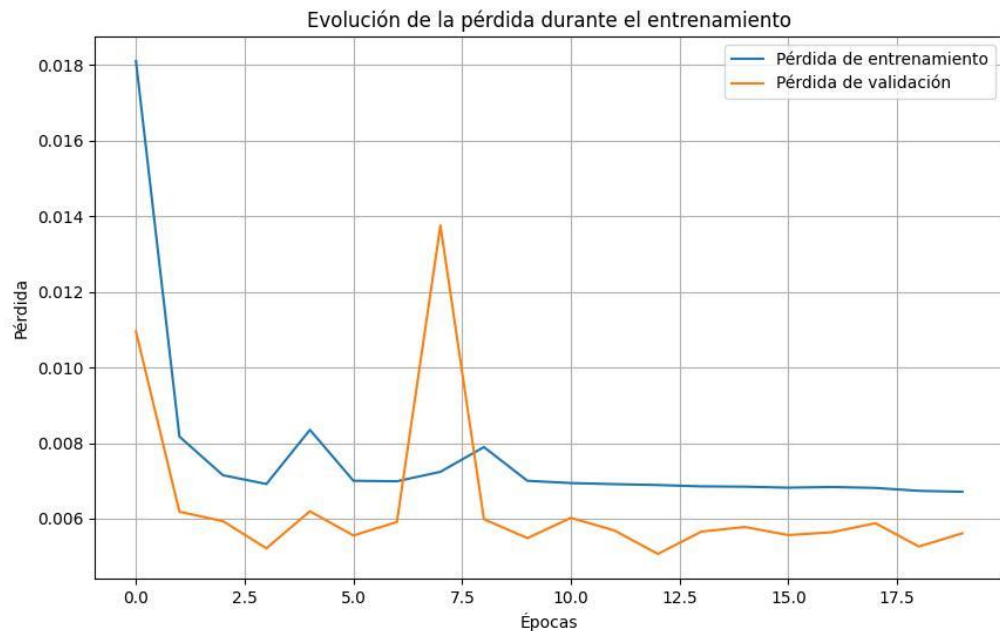
Predicción semanal del Modelo B (LSTM).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

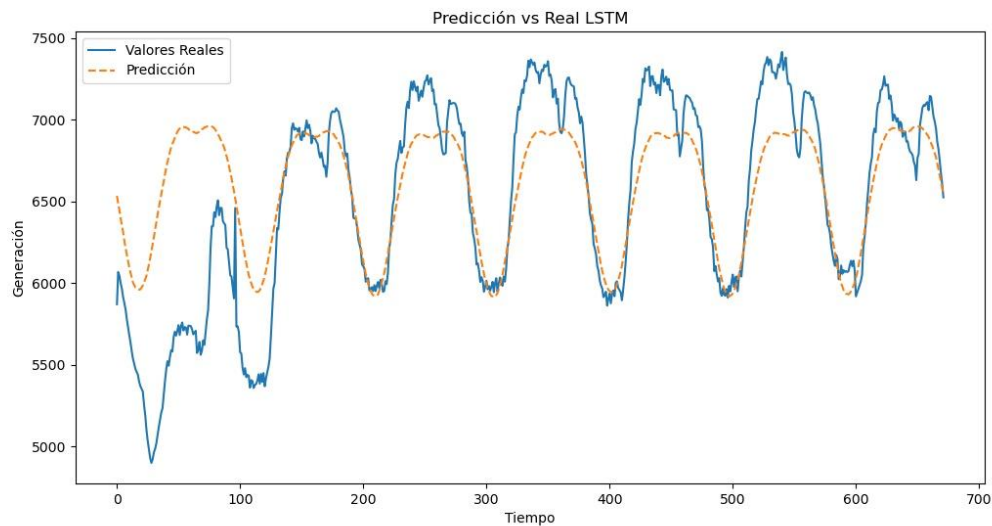
Figura 43

Pérdida de entrenamiento del Modelo C (LSTM).



Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 44
Predicción semanal del Modelo C (LSTM).



Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

La Tabla 11 compara la función de pérdida (Loss) en las fases de entrenamiento y validación, lo que permite evaluar el grado de generalización de cada modelo y determinar la arquitectura más adecuada. Se observa que el modelo A presenta las pérdidas más bajas en ambos conjuntos, lo que sugiere una arquitectura con mejor capacidad de generalización.

Tabla 11
Loss de los modelos con diferentes arquitecturas LSTM.

Modelo	Descripción	Train Loss	Validation Loss
Modelo A	2 capas	0.001315	0.002025
Modelo B	3 capas	0.002749	0.003194
Modelo C	4 capas	0.004541	0.005067

Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

La Tabla 12 sirve para validar la performance de los tres modelos respecto sus métricas de error. Los resultados muestran que una arquitectura más compleja y con más capas no necesariamente mejora los resultados.

Tabla 12

Métricas de error de los modelos con diferentes arquitecturas LSTM.

Modelo	Descripción	MAE	RMSE	MAPE	R Squared
Modelo A	2 capas	180.9068	230.8566	2.8657	0.8610
Modelo B	3 capas	192.8978	283.0747	3.1089	0.7910
Modelo C	4 capas	303.1163	445.3100	4.9365	0.4829

Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Del análisis de los resultados, se concluye que el Modelo A, con dos capas, proporciona la mejor generalización. Además, al tener una menor complejidad arquitectónica, su demanda computacional es reducida, lo que optimiza el tiempo de ejecución sin comprometer la precisión.

Seleccionada la arquitectura óptima, se procede a entrenar una nueva versión del modelo incorporando la variable endógena de potencia ejecutada, con el propósito de evaluar su impacto en la predicción. La Tabla 13 resume las características de entrada (features) de cada modelo, permitiendo comparar su desempeño en términos de error y capacidad predictiva.

Tabla 13

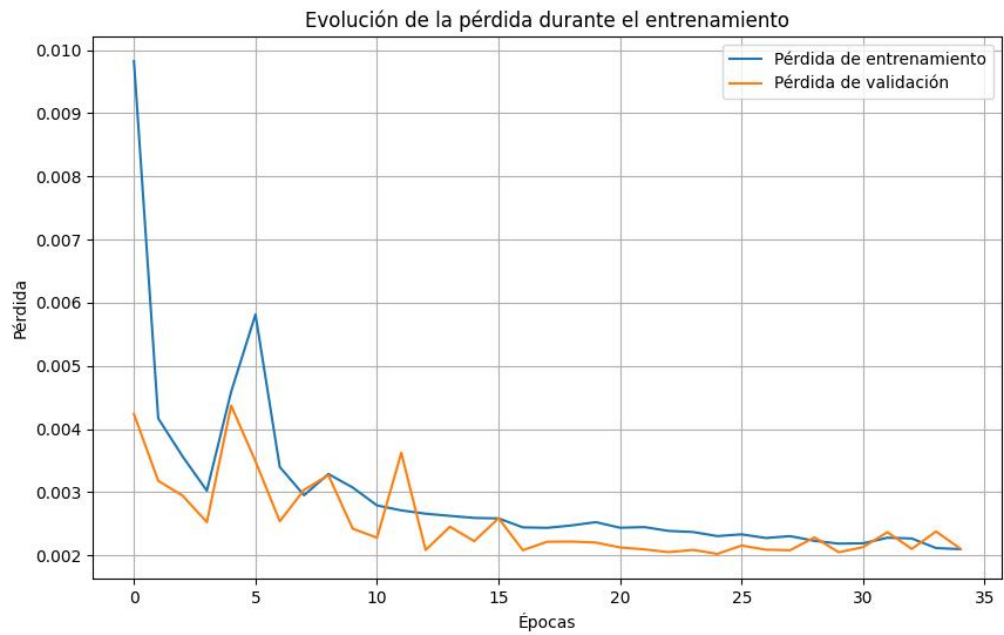
Modelo LSTM con diferentes features

Modelos	Generación	Ejecutado
Modelo 1	X	-
Modelo 2	X	X
Modelo 3	-	X

Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

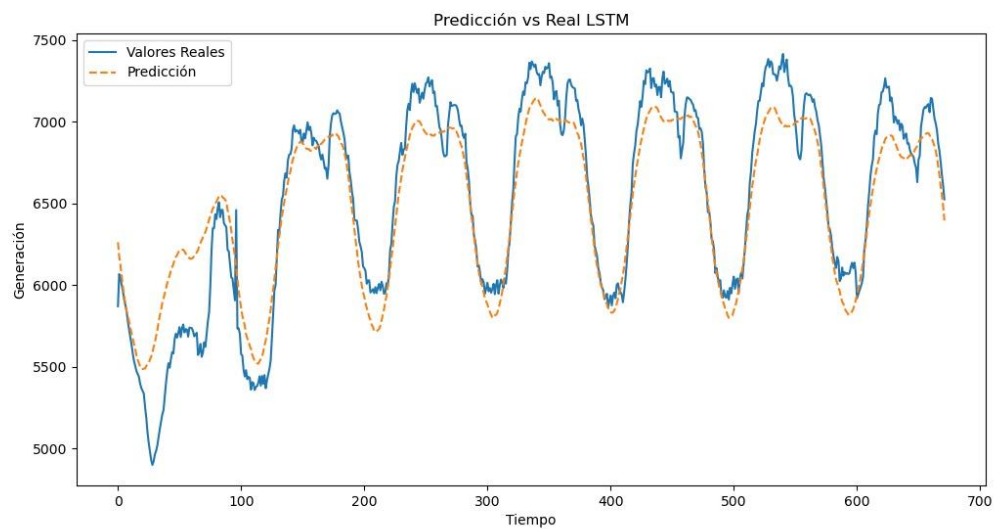
El entrenamiento siguió la misma metodología previamente descrita respetando los colores y características anteriormente mencionadas. Las figuras 45, 48 y 51 muestran la evolución de la función de pérdida (Loss) durante el entrenamiento; las figuras 46, 49 y 52 presentan las predicciones para una semana representativa; y las figuras 47, 50 y 53, el desempeño sobre el conjunto completo de validación. Estas visualizaciones permiten evaluar el comportamiento del modelo en distintos escenarios temporales.

Figura 45
Pérdida de entrenamiento del Modelo 1 (Generación).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

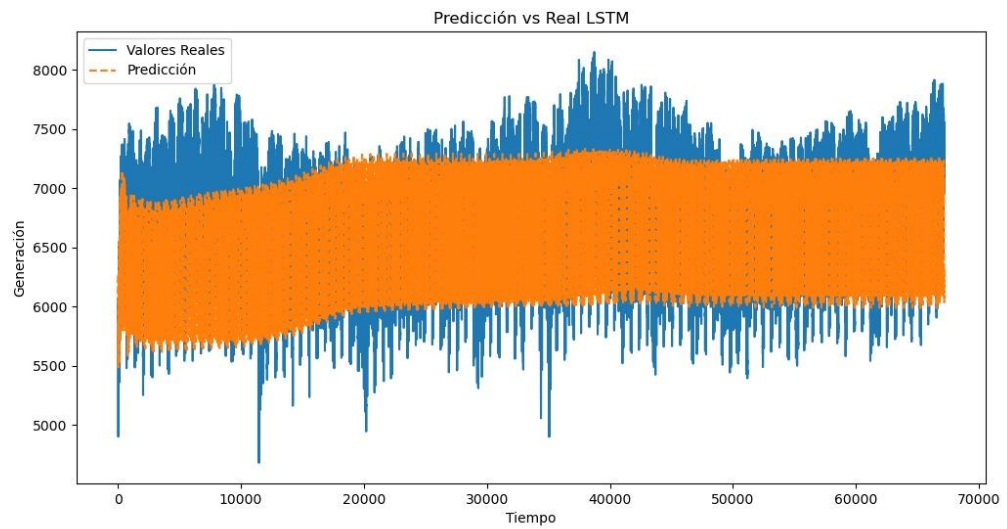
Figura 46
Predicción semanal del Modelo 1 (Generación).



Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 47

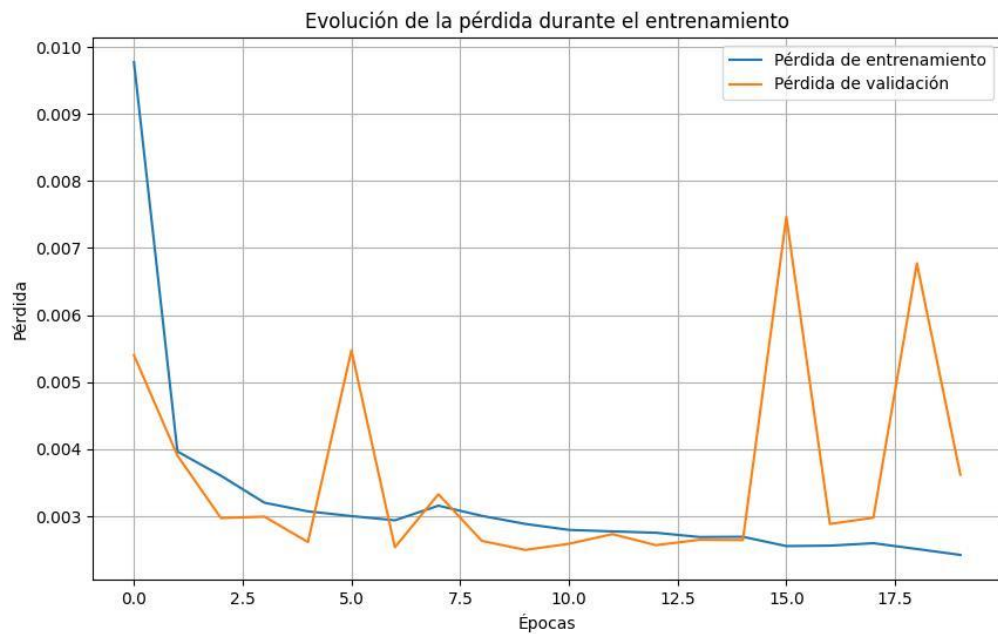
Predicción de validación del Modelo 1 (Generación).



Nota: Elaboración propia en Jupyter Notebook con TensorFlow/Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 48

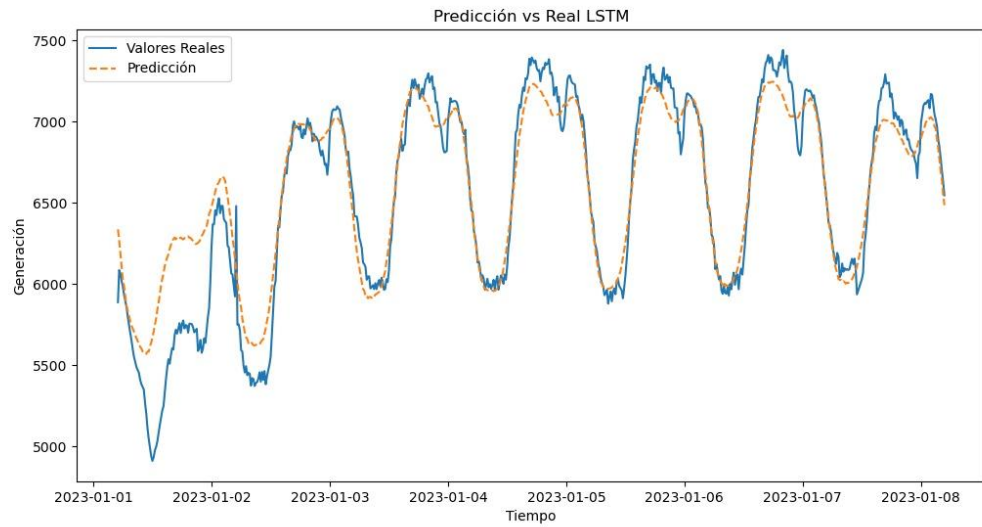
Pérdida de entrenamiento del Modelo 2 (Generación y Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), obtenidos de <https://www.coes.org.pe>

Figura 49

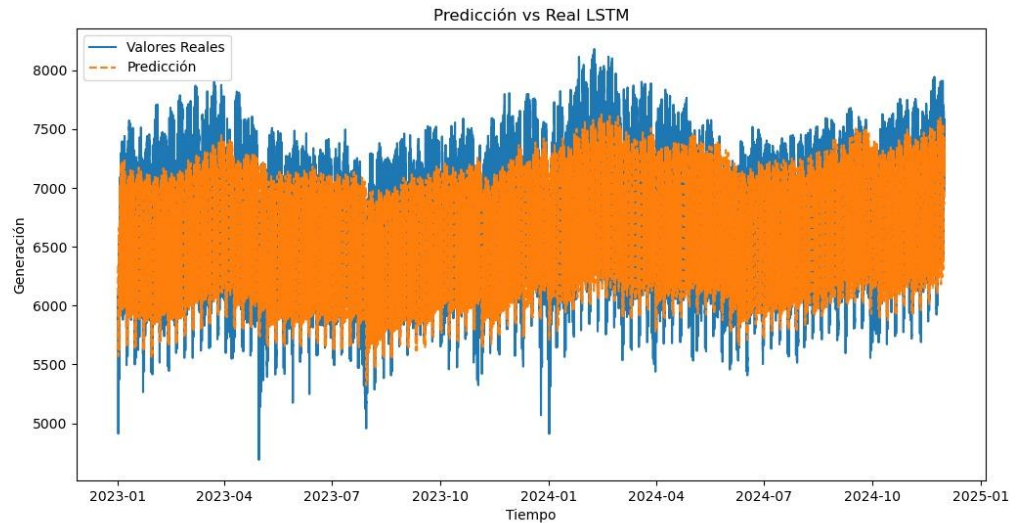
Predicción semanal del Modelo 2 (Generación y Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 50

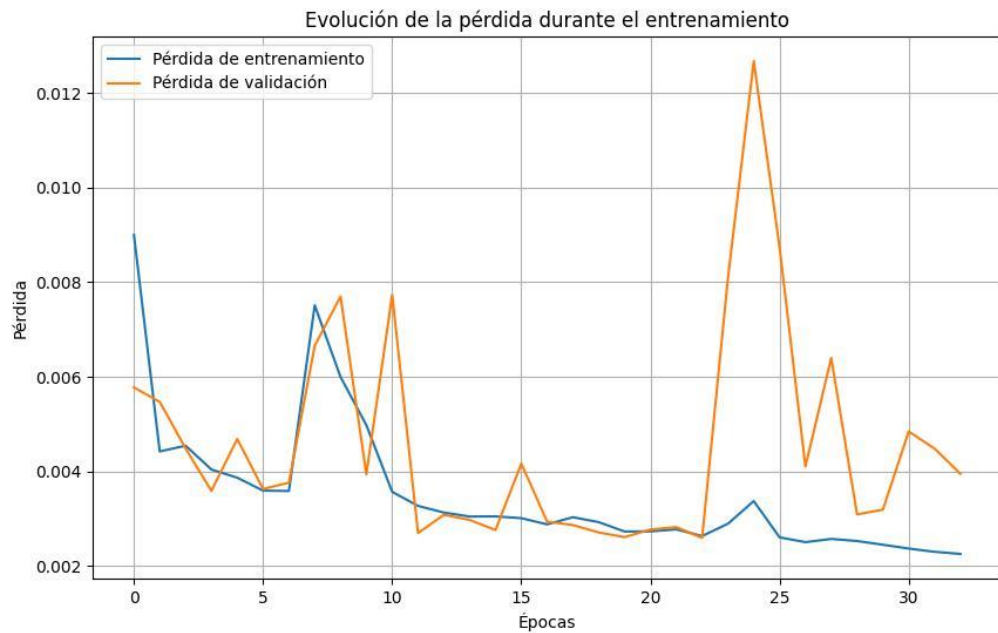
Predicción de validación del Modelo 2 (Generación y Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 51

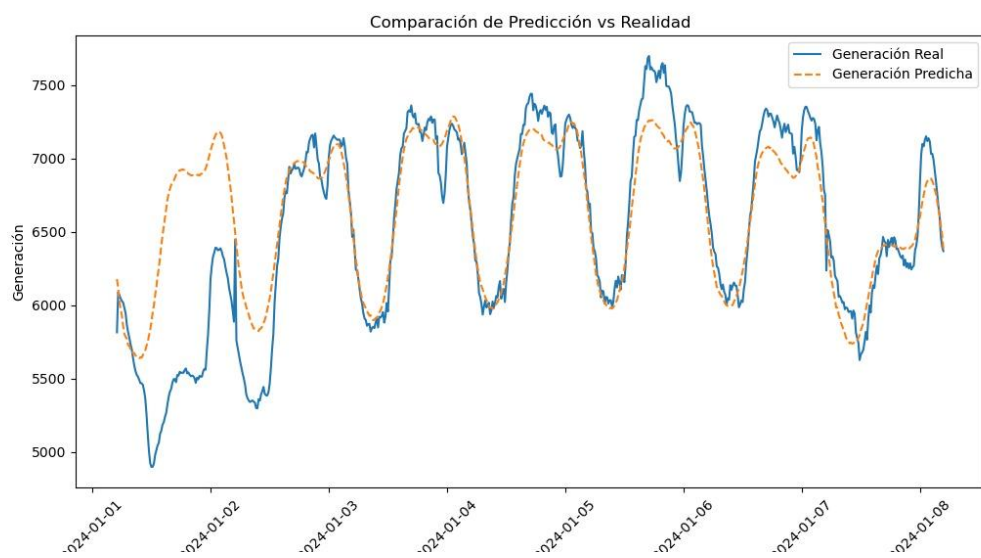
Pérdida de entrenamiento del Modelo 3 (Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

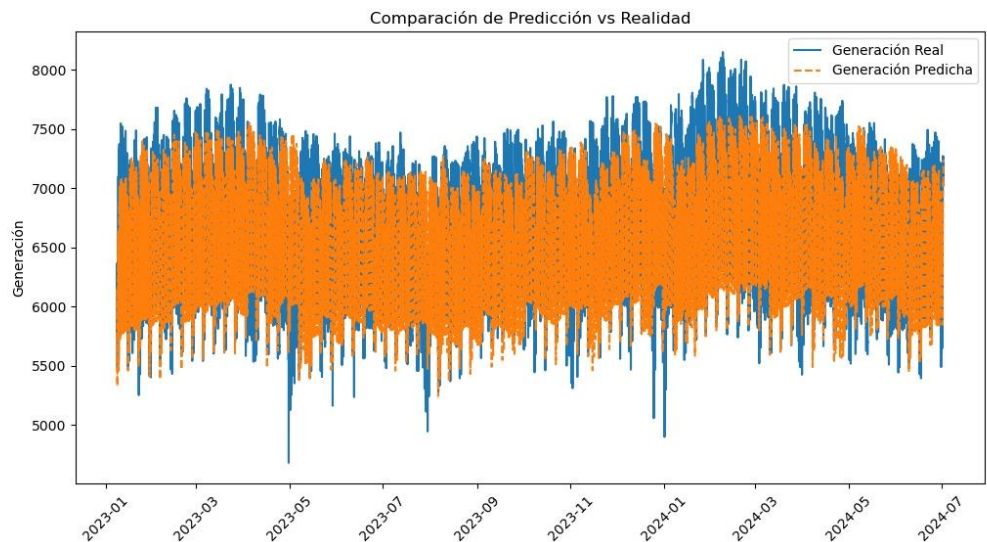
Figura 52

Predicción semanal del Modelo 3 (Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 53
Predicción de validación del Modelo 3 (Ejecutado).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

La Tabla 14 presenta la comparación del desempeño de los modelos considerando diferentes características de entrada (features), con el objetivo de evaluar su impacto en la predicción de la generación de potencia, utilizada como variable objetivo (target).

Tabla 14
Performance de los modelos con diferentes features.

Modelo	Features	MAE	RMSE	MAPE	R Squared
Modelo 1	Generación	180.9068	230.8566	2.8657	0.8610
Modelo 2	Generación, Ejecutado	173.2920	221.9294	2.6104	0.8571
Modelo 3	Ejecutado	256.6423	425.4072	4.2698	0.5950

Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Se observa que el Modelo 3 presenta los mayores errores y un coeficiente de determinación (R^2) significativamente menor en comparación con los otros dos modelos, lo que indica un bajo desempeño en la generalización de los datos. Si bien el Modelo 1 exhibe un R^2 más alto, también presenta errores más elevados. Además, dado que la generación es una variable que se publica con retraso, su uso como única característica de entrada podría generar una propagación del error a lo largo del tiempo. Se experimentó con ambos modelos en el tiempo y se notó que el accuracy tuvo una caída dominante en el modelo A,

corroborando que el modelo B, el cual actualiza datos semanalmente es más robusto que el modelo A, el cual tiene que esperar un mes completo para actualizar información. Por este motivo, se selecciona el Modelo 2, el cual demuestra mayor robustez y una mejor capacidad de generalización.

Definida la arquitectura y las variables de entrada, se procede a evaluar la influencia del tamaño de la ventana de entrada en el rendimiento del modelo. Hasta este punto, el entrenamiento se ha realizado utilizando una ventana de una semana como entrada (feature), para predecir una semana como salida (target). En la Tabla 15, se presentan las diferentes configuraciones probadas con el fin de identificar la opción óptima para la predicción.

Tabla 15

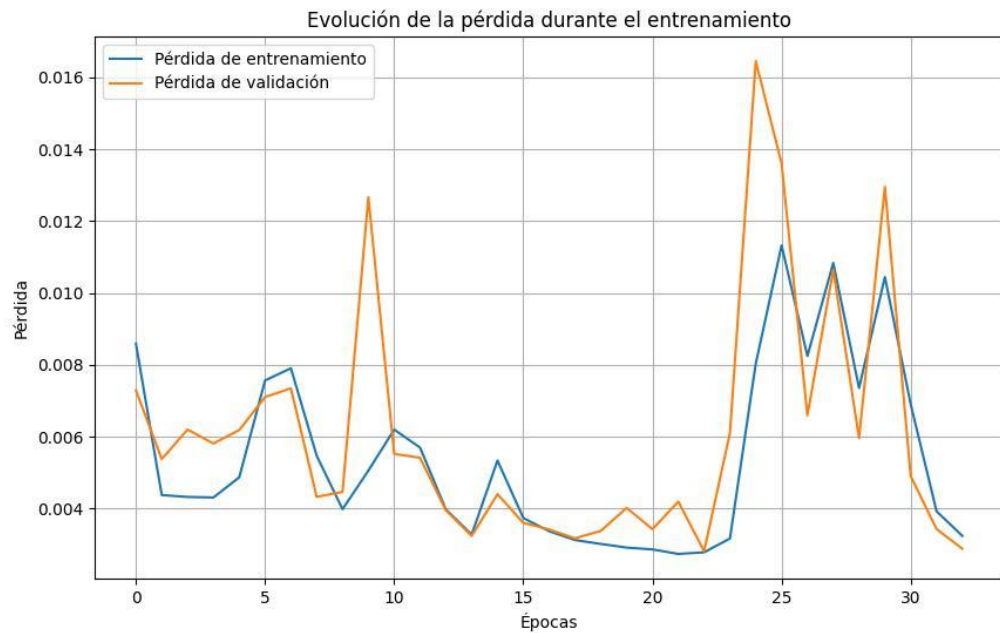
Modelo LSTM con diferentes configuraciones de entrada.

Modelos	Window Size	Prediction Steps
Modelo X	672	672
Modelo Y	1 344	672
Modelo Z	2 688	672

Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

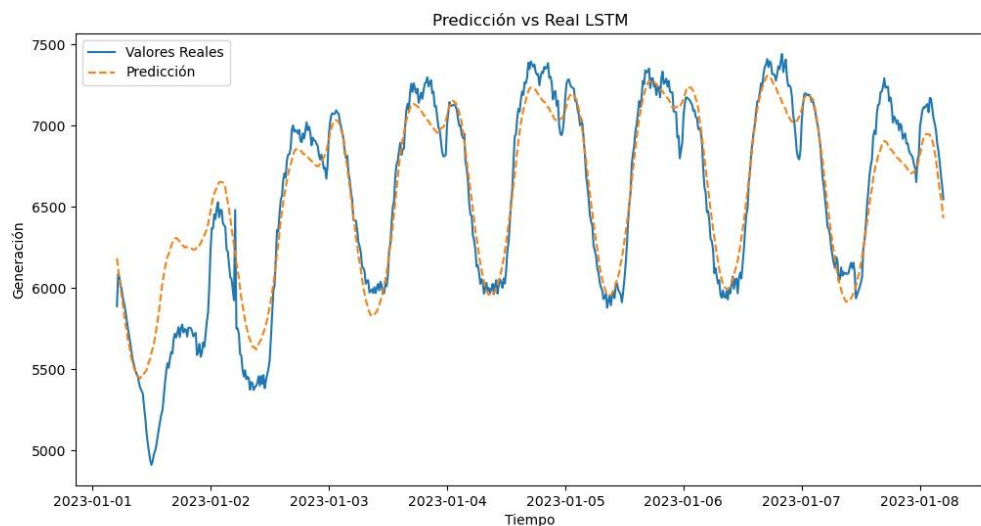
El Modelo X corresponde al Modelo 2, que ya ha sido evaluado y seleccionado previamente. Por lo tanto, el análisis se centrará en la evaluación del Modelo Y y el Modelo Z. En las figuras del 54 al 59, respetando el lineamiento ya mencionado, se muestran las gráficas correspondientes al entrenamiento de ambos modelos, incluyendo la evolución de la función de pérdida (Loss) y sus predicciones en el conjunto de prueba.

Figura 54
Pérdida de entrenamiento del Modelo Y (1344).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

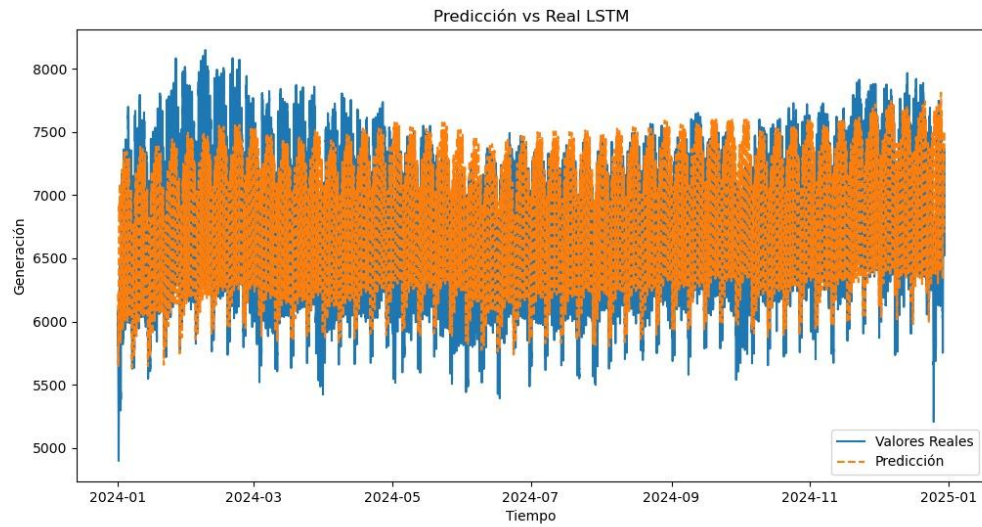
Figura 55
Predicción semanal del Modelo Y (1344).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 56

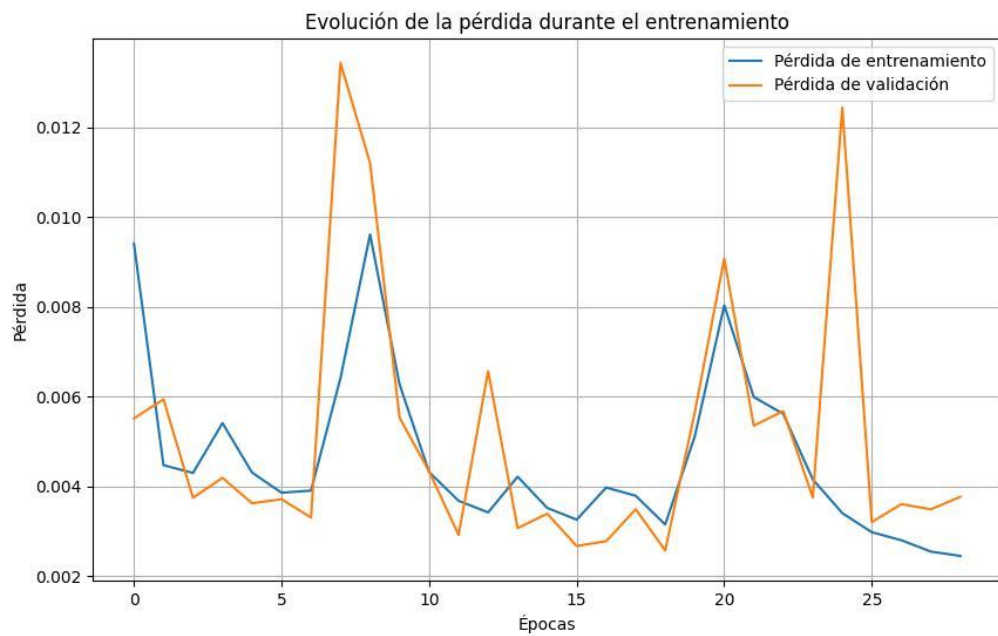
Predicción de validación del Modelo Y (1344).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

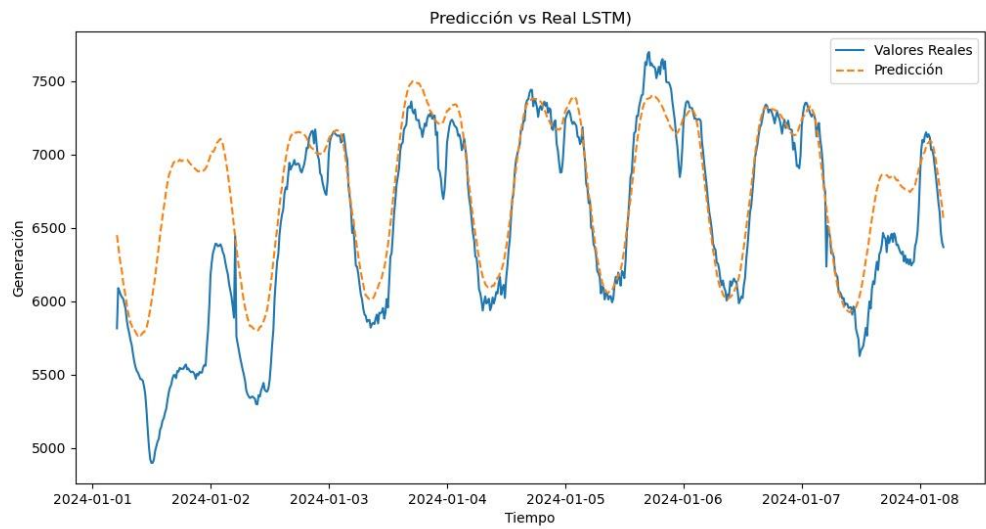
Figura 57

Pérdida de entrenamiento del Modelo Z (2688).



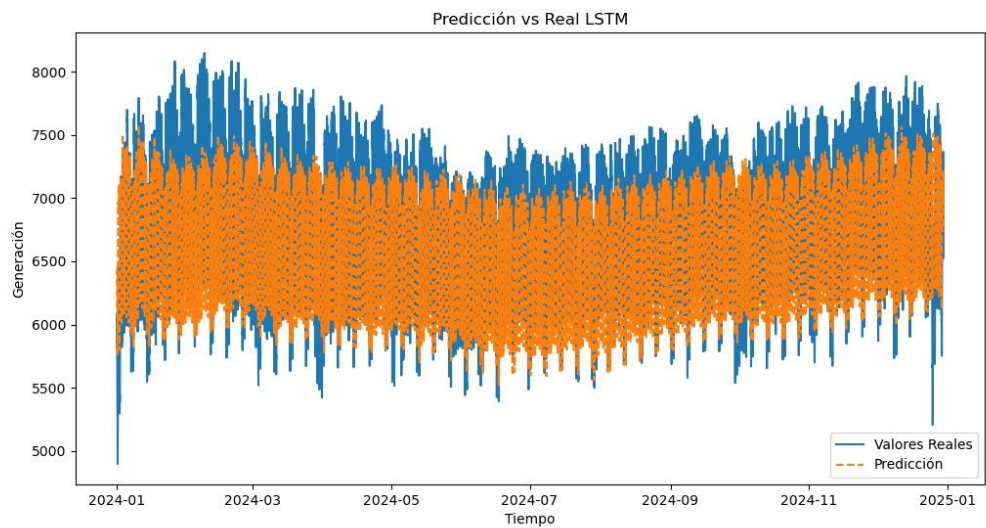
Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 58
Predicción semanal del Modelo Z (2688).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 59
Predicción de validación del Modelo Z (2688).



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

La Tabla 16 presenta la evaluación del desempeño de los modelos utilizando diferentes configuraciones de características de entrada (features). Se observa que los Modelos X e Y exhiben comportamientos muy similares en términos de desempeño, lo que

dificulta la selección de una única opción basándose únicamente en estas características. Por esta razón, ambos modelos se mantienen en el análisis para continuar con una evaluación más profunda.

Tabla 16
Rendimiento del modelo LSTM según tamaño de ventana de entrada.

Modelo	Windows Size	MAE	RMSE	MAPE	R Squared
Modelo X	672	173.2920	221.9294	2.6104	0.8571
Modelo Y	1344	158.4503	223.7927	2.5700	0.8710
Modelo Z	2688	276.5874	447.3781	4.6616	0.5521

Nota: Elaboración propia en Jupyter Notebook con datos del COES (2024). Las métricas reflejan el rendimiento predictivo de modelos LSTM con distintas longitudes de ventana temporal.

Durante el análisis exploratorio se identificó que los feriados y los días de la semana tienen un impacto significativo en la generación de potencia. Dado que estos factores pueden influir en los patrones de consumo y demanda, se decide incorporarlos como variables exógenas en el modelo con el objetivo de mejorar la precisión de las predicciones. La Tabla 17 presenta los modelos que incluirán estas nuevas variables para su evaluación.

Tabla 17
Desempeño de modelos LSTM con distintas combinaciones de variables de entrada.

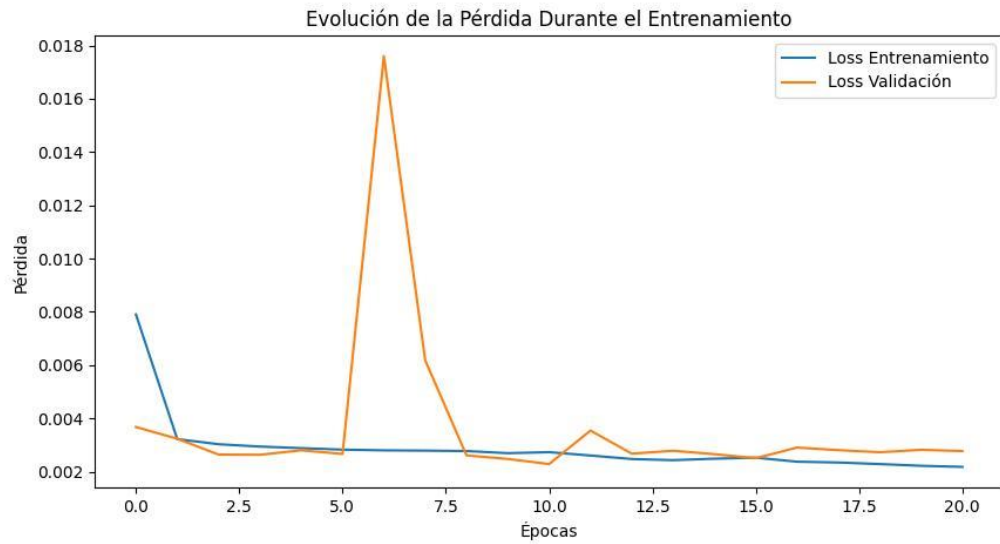
Modelo	Windows Size	Predictions Step	Features	Target
Modelo X	672	672	Generación, Ejecutado	Generación
Modelo Y	1344	672	Generación, Ejecutado	Generación
Modelo XX	672	672	Generación, Ejecutado, Feriados, Días	Generación
Modelo YY	1344	672	Generación, Ejecutado, Feriados, Días	Generación

Nota: Elaboración propia a partir de modelos entrenados en Jupyter Notebook. La variable objetivo fue la potencia generada (Generación), evaluando distintos tamaños de ventana y conjuntos de variables predictoras.

Los siguientes apartados muestran los resultados de los modelos XX y YY, que incorporan las nuevas variables exógenas (feriados y días de la semana), desde la figura 60 hasta la figura 65, evaluando su impacto en la mejora del desempeño predictivo.

Figura 60

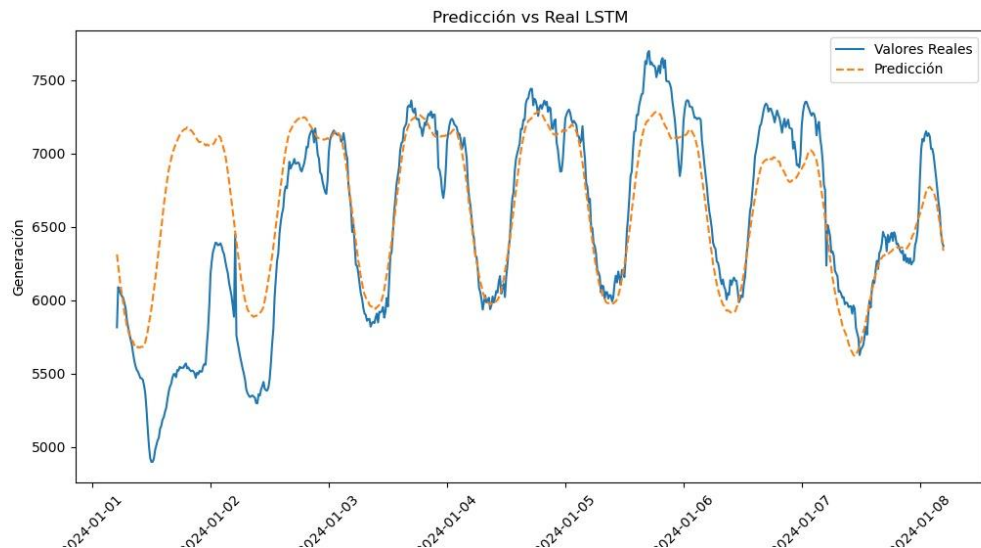
Pérdida de entrenamiento del Modelo XX.



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 61

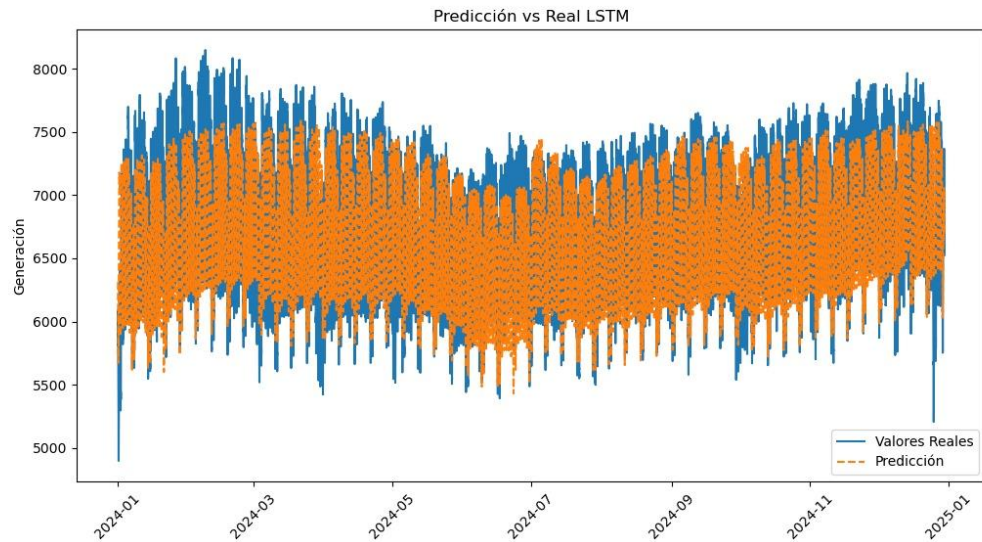
Predicción semanal del Modelo XX.



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 62

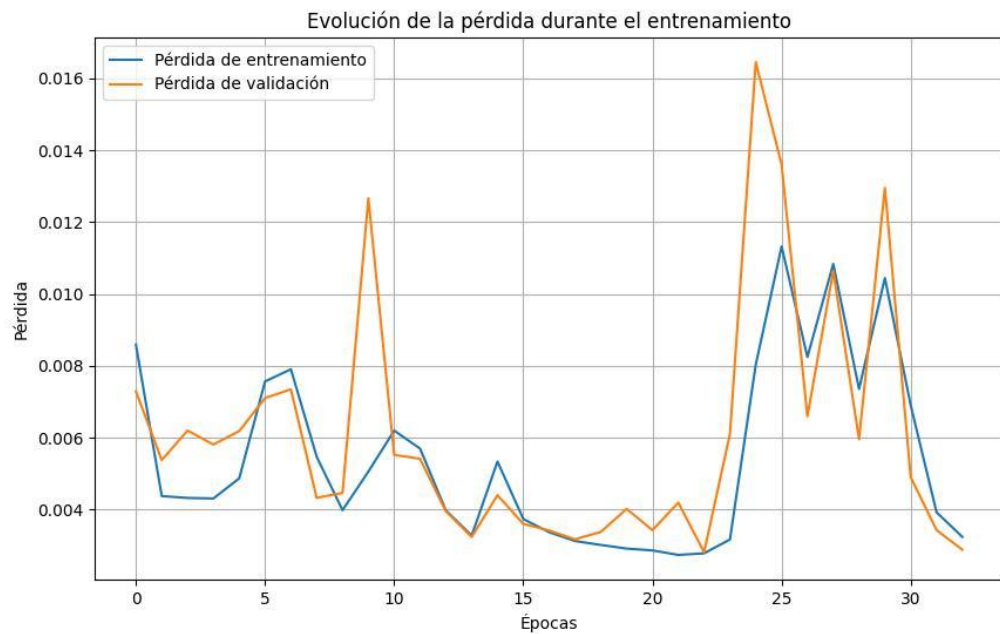
Predicción de validación del Modelo XX.



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

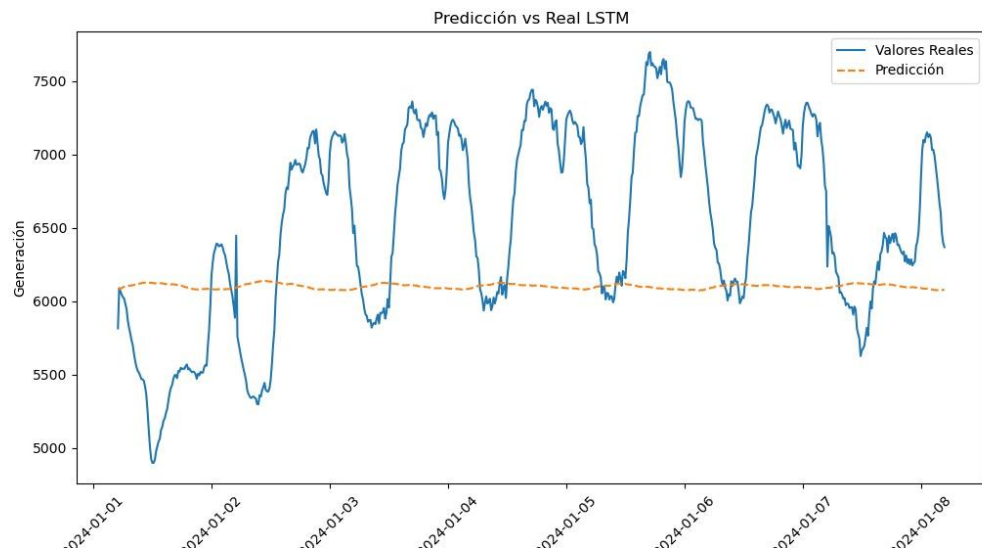
Figura 63

Pérdida de entrenamiento del Modelo YY.



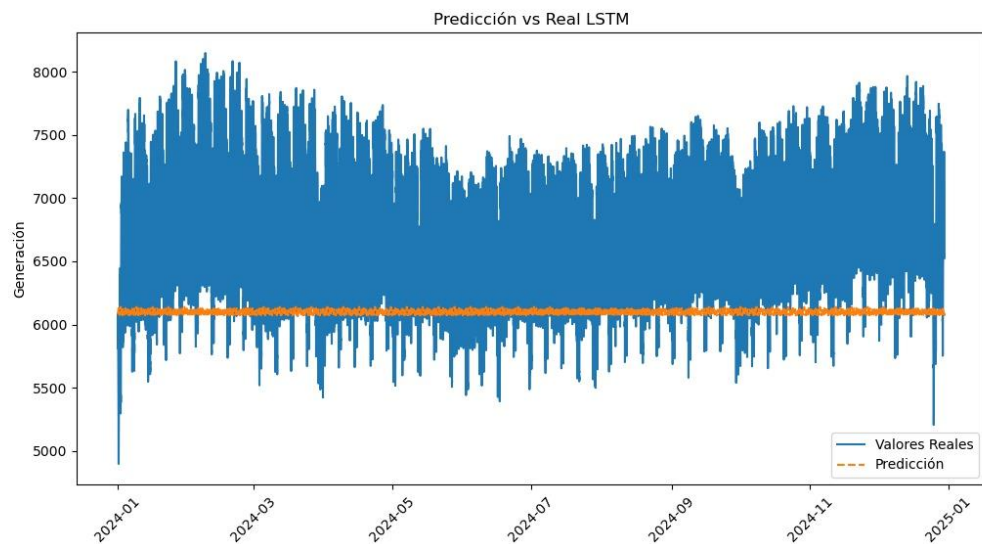
Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 64
Predicción semanal del Modelo YY.



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Figura 65
Predicción de validación del Modelo YY.



Nota: Elaboración propia en Jupyter Notebook con Keras. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Como se observa en las imágenes de la figura 64 y 65, el modelo YY presenta serias deficiencias en términos de generalización. A pesar de múltiples intentos de reentrenamiento para descartar posibles fallas en el proceso de aprendizaje, los resultados obtenidos fueron consistentes, evidenciando una predicción deficiente que no representa adecuadamente el comportamiento del target. En consecuencia, se descarta completamente la evaluación de este modelo YY, y se procede a una comparación detallada entre los modelos X, Y y XX.

Para garantizar un análisis riguroso y alineado con condiciones reales de operación, se simula un escenario de predicción real, en el cual cada domingo se realizan estimaciones utilizando toda la información real disponible hasta la fecha. Dado que la generación se publica mensualmente, el proceso se apoya en predicciones previas para completar los valores faltantes, asegurando así una evaluación integral del desempeño de cada modelo en cada mes del año de prueba. La Tabla 18 presenta el análisis detallado del modelo X, la tabla 19 el análisis del modelo Y, y finalmente la tabla 20 muestra el modelo XX.

Tabla 18

Métricas del desempeño del Modelo X durante el año 2024.

Mes	MAE	RMSE	MAPE	R_Squared
Enero	203.490162	273.437655	2.883627	0.798466
Febrero	198.957702	260.370346	2.746261	0.82845
Marzo	193.845271	264.07148	2.816142	0.814241
Abril	156.529848	232.837831	2.312378	0.827717
Mayo	189.41014	258.313256	2.879485	0.739505
Junio	202.422192	238.060391	3.010732	0.787491
Julio	134.442245	174.972879	2.03868	0.879243
Agosto	177.601668	226.02163	2.623947	0.794462
Setiembre	155.318931	192.637942	2.316477	0.864984
Octubre	209.602825	245.240347	3.020022	0.787006
Noviembre	216.607731	253.440776	3.036421	0.791996
Diciembre	213.421254	309.084297	3.122761	0.714533
	196.401487	249.340562	2.8478135	0.796464

Nota: Elaboración propia a partir de las predicciones mensuales del Modelo X. Las métricas corresponden al conjunto de validación. Cálculos realizados en Jupyter Notebook.

Tabla 19*Métricas del performance del Modelo Y durante todo el 2024.*

Mes	MAE	RMSE	MAPE	R_Squared
Enero	184.042898	228.994581	2.643561	0.858655
Febrero	223.642702	273.481488	3.187001	0.810738
Marzo	206.229589	288.392045	3.056684	0.778449
Abril	197.845711	284.983984	2.979202	0.741907
Mayo	220.736993	276.197411	3.379991	0.702186
Junio	152.668002	199.728796	2.355291	0.850416
Julio	173.39018	239.814637	2.690657	0.773159
Agosto	170.928169	229.12432	2.588524	0.78878
Setiembre	251.659472	319.37987	3.795683	0.628878
Octubre	171.773125	234.637667	2.54058	0.805025
Noviembre	169.183598	209.651416	2.444727	0.857664
Diciembre	216.907793	339.821247	3.274427	0.654933
	190.944305	256.648063	2.8349295	0.7836145

Nota: Elaboración propia a partir de las predicciones mensuales del Modelo Y. Las métricas corresponden al conjunto de validación. Cálculos realizados en Jupyter Notebook.

Tabla 20*Métricas del performance del Modelo XX durante todo el 2024.*

Mes	MAE	RMSE	MAPE	R_Squared
Enero	181.100429	227.657617	2.550502	0.8603
Febrero	200.773416	251.635941	2.77812	0.839766
Marzo	186.89228	248.356216	2.727764	0.835693
Abril	159.429421	234.072734	2.353959	0.825884
Mayo	112.510293	141.906908	1.681376	0.921384
Junio	201.2876	238.96541	3.006794	0.785872
Julio	160.123581	197.24107	2.408057	0.846551
Agosto	139.430149	173.268411	2.047662	0.87921
Setiembre	138.854286	174.215385	2.058131	0.889573
Octubre	182.217527	219.31964	2.632791	0.829651
Noviembre	208.702543	245.410067	2.923286	0.804969
Diciembre	207.851013	310.546566	3.057645	0.711825
	181.658978	230.865176	2.5916465	0.8377295

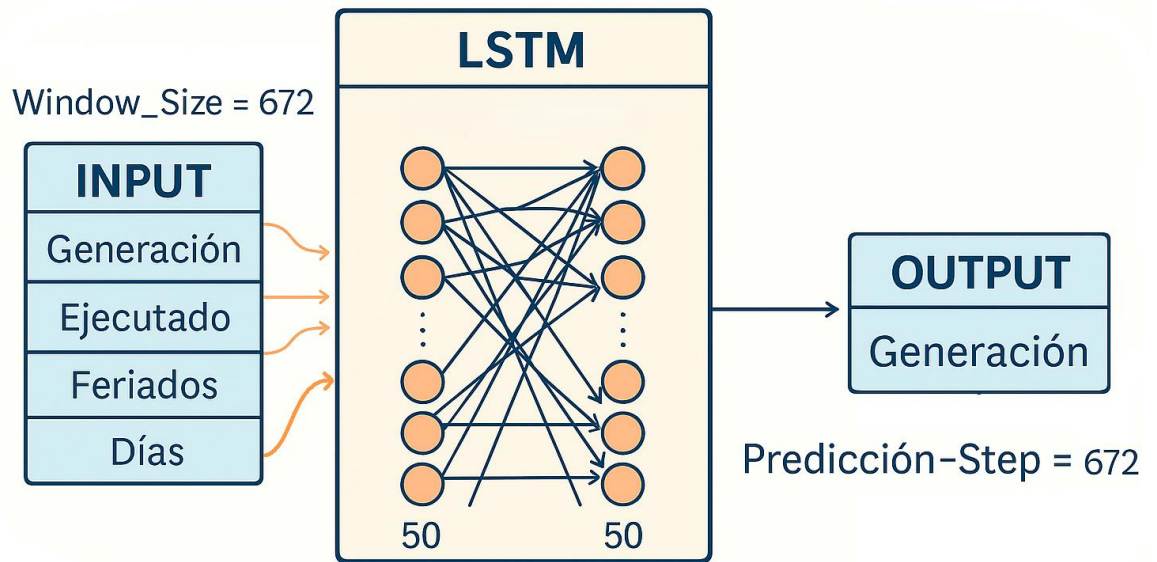
Nota: Elaboración propia a partir de las predicciones mensuales del Modelo XX. Las métricas corresponden al conjunto de validación. Cálculos realizados en Jupyter Notebook.

Dado su desempeño superior, el modelo XX se selecciona como el definitivo para las predicciones en esta investigación. La figura 66 muestra detalladamente el diseño del modelo LSTM seleccionado. Son 4 las variables de ingreso (Generación, Ejecutado, Feriados y Días de la semana) presentadas a lado izquierdo con un tamaño de entrada de

672 valores, equivalente a una semana. El modelo LSTM presentado al medio y compuesto por dos capas de 50 neuronas, con una variable de salida, 'Generación', de tamaño de 672 valores, que representan la siguiente semana la cual se quiere predecir.

Figura 66

Diagrama del modelo LSTM seleccionado, incluyendo variables de entrada y salida.



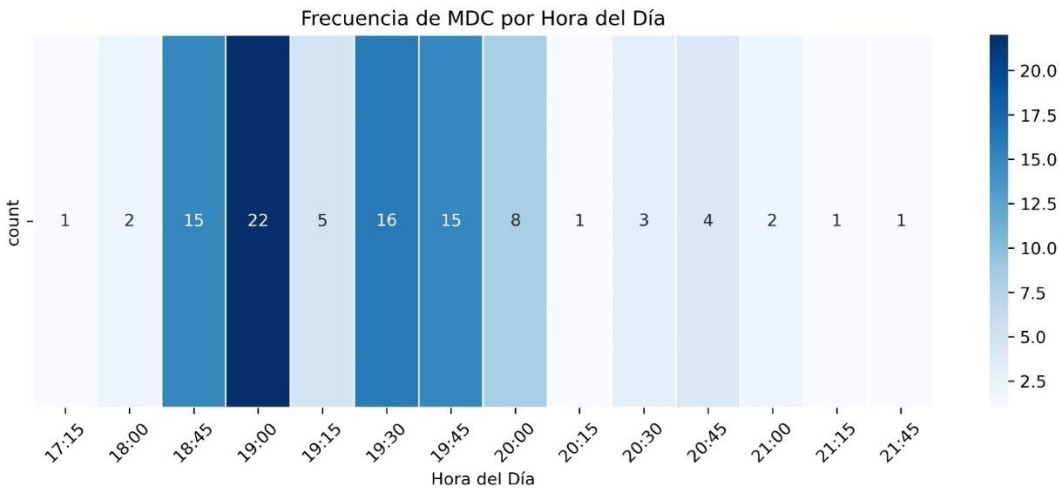
Nota: Elaboración propia con asistencia de ChatGPT y edición en Windows.

3.3 Rango horario de la MDC

Las horas exactas en las que ocurre la MDC siguen presentando cierto nivel de variabilidad. Por esta razón, se define un rango horario representativo que encapsule la mayoría de los eventos de MDC. La figura 67 muestra la distribución de las horas más frecuentes en las que se registró la MDC, observándose una mayor concentración entre las 18:45 h y las 20:45 h. Este rango de dos horas abarca el 92.7 % de las ocurrencias de MDC, lo que permite un enfoque más efectivo para la implementación de estrategias de reducción de demanda.

Figura 67

Frecuencia por hora del día de la Máxima Demanda Coincidente (2017–2024).



Nota: Elaboración propia a partir de datos históricos de la MDC procesados en Jupyter Notebook. Datos del Comité de Operación Económica del SEIN (COES, 2024), disponibles en <https://www.coes.org.pe>

Capítulo IV. Análisis de resultados

El presente capítulo expone los principales hallazgos obtenidos a partir de la implementación del modelo de aprendizaje profundo seleccionado en el capítulo anterior (Modelo XX) para la predicción de la Máxima Demanda Coincidente (MDC) en el Sistema Eléctrico Interconectado Nacional (SEIN).

Se realizan predicciones semanales para validar la precisión del modelo en la estimación de la MDC. Aunque la MDC ocurre una sola vez al mes, la predicción semanal permite detectar con mayor resolución los picos de demanda, facilitando así la aplicación de estrategias como peak shaving y load shifting. La cantidad de días en los que estas estrategias serían implementadas dependerá del perfil de consumo específico del usuario libre (UL).

Las predicciones semanales se someten a un análisis de calidad con respecto a la estimación de la MDC. La Tabla 21 muestra los resultados obtenidos de la predicción de todo el 2024, extrayendo el valor máximo semanal como referencia para la evaluación final.

Tabla 21

Métricas del performance de la predicción del Modelo XX durante todo el 2024.

Mes	Semana	Fecha	Hora	Potencia	MDC
Enero	1	1/10/2024	18:00	7312.023	0
Enero	2	1/18/2024	20:00	7243.2915	0
Enero	3	1/25/2024	20:00	7328.717	0
Enero	4	1/31/2024	18:30	7420.769	0
Febrero	1	2/7/2024	18:30	7488.622	1
Febrero	2	2/14/2024	18:30	7517.13	0
Febrero	3	2/21/2024	18:30	7509.3267	0
Febrero	4	2/28/2024	18:30	7467.41	0
Marzo	1	3/6/2024	18:30	7313.6997	0
Marzo	2	3/13/2024	18:30	7296.5723	0
Marzo	3	3/20/2024	18:30	7309.714	0
Marzo	4	3/27/2024	18:30	7351.2896	0
Abril	1	4/3/2024	19:45	7393.342	0
Abril	2	4/10/2024	18:30	7330.376	0
Abril	3	4/16/2024	19:45	7372.777	0
Abril	4	4/23/2024	19:45	7328.1904	0
Abril	5	4/30/2024	19:45	7391.3496	0
Mayo	1	5/9/2024	20:00	7296.4375	0

Mes	Semana	Fecha	Hora	Potencia	MDC
Mayo	2	5/16/2024	20:00	7256.27	1
Mayo	3	5/23/2024	20:00	7247.8726	0
Mayo	4	5/28/2024	19:45	7174.473	0
Junio	1	6/6/2024	20:00	7036.0723	0
Junio	2	6/13/2024	20:00	7092.552	0
Junio	3	6/20/2024	20:00	7070.625	0
Junio	4	6/27/2024	20:00	7107.822	0
Julio	1	7/3/2024	20:15	7440.074	0
Julio	2	7/11/2024	20:00	7260.156	0
Julio	3	7/18/2024	20:00	7203.0776	0
Julio	4	7/25/2024	20:00	7163.7837	0
Agosto	1	8/1/2024	20:00	7194.8286	0
Agosto	2	8/8/2024	20:00	7298.8105	0
Agosto	3	8/15/2024	20:00	7279.143	0
Agosto	4	8/22/2024	20:00	7241.662	0
Agosto	5	8/29/2024	20:00	7302.7637	0
Setiembre	1	9/5/2024	20:00	7315.7544	0
Setiembre	2	9/12/2024	20:00	7408.4346	0
Setiembre	3	9/17/2024	19:45	7435.828	0
Setiembre	4	9/24/2024	19:45	7424.4	0
Octubre	1	10/3/2024	19:45	7306.392	0
Octubre	2	10/8/2024	19:45	7222.959	0
Octubre	3	10/17/2024	20:00	7284.272	0
Octubre	4	10/24/2024	20:00	7305.783	1
Octubre	5	10/29/2024	19:45	7392.528	0
Noviembre	1	11/6/2024	19:45	7422.968	0
Noviembre	2	11/14/2024	20:00	7382.0176	0
Noviembre	3	11/19/2024	19:45	7421.8022	1
Noviembre	4	11/26/2024	19:45	7503.915	0
Diciembre	1	12/4/2024	18:30	7555.761	0
Diciembre	2	12/10/2024	19:45	7547.9775	0
Diciembre	3	12/17/2024	19:45	7550.369	0
Diciembre	4	12/24/2024	19:45	7534.6016	0

Nota: Las celdas resaltadas indican las fechas en las que se detectó una ocurrencia de MDC

La columna MDC es 1 si la predicción del día concuerda con la Máxima Demanda Coincidente del mes, caso contrario vale 0. La validación de esta información se realizó con las 12 MDC registradas en el año 2024, como se detalla en la Tabla 22.

Tabla 22*Fecha y hora del MDC mensual del 2024.*

Fecha	Hora	Potencia
26/01/2024	19:45	7632.53124
07/02/2024	20:30	7761.9886
19/03/2024	19:00	7550.30693
24/04/2024	18:45	7548.01572
16/05/2024	18:45	7431.31068
21/06/2024	19:00	7347.19681
23/07/2024	20:45	7416.906
20/08/2024	18:45	7447.275
10/09/2024	18:45	7581.449
24/10/2024	19:00	7583.628
19/11/2024	20:00	7794.008
03/12/2024	19:30	7698.94

Nota: MDC corresponde al momento de máxima demanda coincidente registrada cada mes.

Aunque las predicciones no coinciden exactamente en las horas correspondientes, se evalúan dentro del rango horario de mayor incidencia seleccionado (18:45 h a 20:45 h), logrando coincidencias exactas en cuatro de las fechas reales de MDC. Con el fin de mejorar los resultados, se incrementa la frecuencia de días por semana de aplicación de las estrategias de Peak Shaving y Load Shifting. La Tabla 23 presenta los resultados coincidentes al aplicar las estrategias dos veces por semana, mientras que la Tabla 24 expone el análisis realizado con tres días semanales.

Tabla 23*Fecha del MDC mensual del 2024 (2 días semanales).*

Fecha (Real)	Fecha (Predicción)	MDC
26/01/2024	31/01/2024	0
07/02/2024	07/02/2024	1
19/03/2024	27/03/2024	0
24/04/2024	03/04/2024	0
16/05/2024	16/05/2024	1
21/06/2024	27/06/2024	0
23/07/2024	23/07/2024	1
20/08/2024	20/08/2024	1
10/09/2024	10/09/2024	1
24/10/2024	24/10/2024	1
19/11/2024	19/11/2024	1
03/12/2024	03/12/2024	1

Nota: El valor "1" indica coincidencia entre la predicción y la fecha real del MDC; "0" indica diferencia.

Tabla 24*Fecha del MDC mensual del 2024 (3 días semanales).*

Fecha (Real)	Fecha (Predicción)	MDC
26/01/2024	31/01/2024	0
07/02/2024	07/02/2024	1
19/03/2024	19/03/2024	1
24/04/2024	24/04/2024	1
16/05/2024	16/05/2024	1
21/06/2024	27/06/2024	0
23/07/2024	23/07/2024	1
20/08/2024	20/08/2024	1
10/09/2024	10/09/2024	1
24/10/2024	24/10/2024	1
19/11/2024	19/11/2024	1
03/12/2024	03/12/2024	1

Nota: El valor “1” indica coincidencia entre la predicción y la fecha real del MDC; “0” indica diferencia

El aumento en la frecuencia semanal de predicción mejora la precisión del modelo, alcanzando 10 coincidencias con los valores reales de MDC al aplicar la estrategia tres veces por semana. La Tabla 25 resume estos resultados.

Tabla 25*MDC acertado según la cantidad de días de la semana aplicados.*

Cantidad de Días	MDC Acertado
1	4
2	8
3	10

Nota: El valor indica la cantidad de meses en los que se acertó la fecha del MDC utilizando el modelo con 1, 2 o 3 días de predicción semanales

La precisión del modelo mostró una mejora progresiva a medida que se incrementó la frecuencia semanal de aplicación de las estrategias predictivas, alcanzando hasta 10 coincidencias con los valores reales de la MDC cuando se aplicó la predicción tres veces por semana. La elección del número de días para ejecutar estas estrategias dependerá del perfil de consumo específico y de la estrategia operativa adoptada por cada Usuario Libre.

Con el objetivo de dimensionar el impacto potencial del modelo predictivo sobre los costos de un Usuario Libre, se presenta un ejemplo de aplicación. Cabe resaltar que los valores utilizados en este análisis son estimaciones referenciales, ya que muchos datos

operativos son de carácter confidencial y se gestionan de forma exclusiva entre el Usuario Libre y su proveedor energético. Asimismo, algunos valores pueden variar en función de la tecnología disponible para cada usuario, por lo que los resultados deben interpretarse como una aproximación y no como valores generalizables a todo contexto. Se recomienda realizar un análisis específico para cada caso particular.

En el presente ejemplo, se considera un Usuario Libre con una demanda coincidente de potencia de $P=290$ MW y una demanda energética anual de $E=206,552$ MWh, cifras que se alinean con los niveles de consumo de los principales usuarios libres del Perú según los registros del COES para el año 2024. La Tabla 26 detalla los costos energéticos asumidos para este análisis. Estos valores son referenciales y pueden variar según las condiciones contractuales, dado que los usuarios libres poseen capacidad de negociación con las empresas generadoras y distribuidoras. Además, se asume que el usuario cuenta con un sistema de almacenamiento de energía (BESS) de 20 MW, capacidad equivalente a la de los sistemas más potentes instalados en el país hasta el año 2024.

Tabla 26
Costos energéticos detallados de un Usuario Libre.

Descripción	Costo	Unidades	Símbolo
Costo de Potencia Firme	6	USD\$ / KW-mes	C_p
Costo de Energía Contratada	32	USD\$ / MWh	C_E
Costo de Peaje por Conexión al Sistema Principal	25	USD \$/MW-mes	C_{SPT}
Costo de Peaje Secundario por Transmisión	4	USD \$/MWh	C_{SST}
Costos Complementarios	2	USD \$/MWh	C_c

Nota: Costos extraídos de los contratos de usuarios libres de OSINERGMIN.

Para el cálculo de los costos energéticos se empleó una adaptación de la ecuación general de facturación, la cual permite integrar los diferentes componentes tarifarios en una sola expresión unificada en dólares, evitando así la separación entre montos expresados en distintas monedas. Esta formulación facilita el análisis económico y permite una comparación clara entre escenarios.

$$Costo_{Mensual} = Costo_{Potencia} + Costo_{Energía} + Costo_{Peaje} + Costos_{Complementarios}$$

$$Costo_{Mensual} = (P * C_p) + (E * C_E) + (P * C_{SPT} + E * C_{SST}) + (E * C_c)$$

La Tabla 27 presenta una comparación entre los costos operativos anuales bajo condiciones normales y aquellos obtenidos al aplicar la estrategia de reducción de demanda máxima utilizando un sistema BESS de 20 MW. Como se observa, se genera un ahorro mensual estimado de \$120 500 (Ciento veinte mil quinientos), directamente asociado a la tecnología disponible para implementar estrategias de load shifting. En este caso, el ahorro depende de la potencia del sistema BESS; variaciones en su capacidad afectarían proporcionalmente el resultado económico. Dado que el modelo logró predecir correctamente la MDC en 10 de los 12 meses del año 2024, el ahorro anual proyectado asciende a aproximadamente \$1 205 000 (Un millón doscientos cinco mil.), lo que demuestra un impacto económico significativo y realista en la operación de usuarios libres de gran escala.

Tabla 27

Tabla comparativa de costos usando y sin usar el modelo de predicción MDC en un usuario libre durante el 2024.

	Potencia	Energía	Peajes	Otros	Total
Sin Load Shifting	1 740 000	6 609 664	833 458	413 104	9 596 226
Con Load Shifting	1 620 000	6 609 664	832 958	413 104	9 475 726

Nota: Costos comparativos en dólares americanos (USD).

Los resultados obtenidos validan que los modelos de Machine Learning pueden ofrecer una prognosis efectiva que habilite a los usuarios a implementar estrategias como peak shaving y load shifting, con el objetivo de optimizar sus costos energéticos

Conclusiones

El desarrollo de esta investigación permitió construir una herramienta predictiva efectiva basada en redes neuronales LSTM, capaz de anticipar con alta precisión los momentos de la Máxima Demanda Coincidente (MDC) en el Sistema Eléctrico Interconectado Nacional (SEIN). Este resultado valida la viabilidad de aplicar modelos de Machine Learning para mejorar la prognosis de la generación de potencia y optimizar la gestión energética de los usuarios libres.

Uno de los principales impactos del modelo radica en su contribución a la optimización del consumo energético. La capacidad predictiva alcanzada facilita la implementación de estrategias como Peak Shaving y Load Shifting, permitiendo reducir costos operativos asociados a la potencia contratada y a los peajes de transmisión, al anticipar los días con mayor probabilidad de ocurrencia de la MDC. En términos de rendimiento, el modelo logró predecir correctamente la MDC en 10 de los 12 meses evaluados durante el año 2024, lo que representa un desempeño altamente satisfactorio. Este resultado se traduce, según el ejemplo de aplicación desarrollado, en un ahorro anual estimado de 1.2 millones de dólares anuales, evidenciando un impacto económico tangible y relevante.

Asimismo, se identificó que el 92.7 % de las MDC se concentran en un intervalo reducido de solo dos horas, entre las 18:45 y las 20:45 horas, dentro del rango de punta definido por el SEIN. Este hallazgo sugiere que, en etapas iniciales, es más eficiente centrar los esfuerzos del modelo en predecir el día de ocurrencia de la MDC, sin necesidad de una precisión horaria exacta, optimizando de esta forma la estrategia de reducción de costos.

Otro hallazgo clave fue la confirmación de la influencia significativa de variables exógenas, como feriados y días de la semana, en la demanda eléctrica. Su inclusión mejora la curva de predicción y permite capturar patrones relevantes del comportamiento de la generación de potencia. Por tanto, se recomienda incorporarlas en futuros modelos para incrementar la precisión.

No obstante, el modelo presenta limitaciones al predecir valores extremos, especialmente en presencia de picos o valles abruptos. Estas dificultades podrían atribuirse a factores no modelados o a deficiencias en la calidad de los datos disponibles. Asimismo, los días festivos introdujeron desviaciones significativas, lo que resalta la importancia de considerar variables adicionales y mejorar la calidad del entrenamiento en estos escenarios atípicos. En esa misma línea, se identificaron inconsistencias en ciertas fuentes de datos del COES, afectando la confiabilidad del modelo. Por ello, es fundamental optimizar la calidad y validación de los datos para aumentar la robustez del sistema predictivo.

Finalmente, la aplicabilidad práctica del modelo en la planificación de la demanda eléctrica demuestra un alto potencial. Su integración con sistemas de gestión de energía (EMS), especialmente en contextos que cuentan con almacenamiento energético (BESS), permitiría automatizar estrategias de gestión de demanda con alto grado de eficiencia. En ese sentido, se concluye que la predicción semanal de la demanda, al identificar con anticipación los días críticos, permite a los usuarios libres tomar decisiones estratégicas de contratación y consumo, contribuyendo directamente a la reducción de costos y al uso más eficiente de los recursos energéticos disponibles.

Recomendaciones

- Integración de arquitecturas híbridas: Se recomienda explorar modelos combinados que integren redes LSTM con redes convolucionales (CNN) o métodos estadísticos tradicionales, con el fin de mejorar la precisión en la predicción de la MDC a través de la captura de patrones tanto temporales como espaciales.
- Incorporación de variables exógenas relevantes: Se sugiere ampliar el conjunto de variables de entrada incluyendo factores meteorológicos, eventos socioeconómicos, feriados, y posibles cambios estructurales en el SEIN, con el objetivo de capturar dinámicas externas que influyen directamente en la demanda eléctrica.
- Tratamiento avanzado de datos atípicos y calidad de datos: Es fundamental implementar técnicas robustas para la detección y manejo de outliers, así como establecer procesos de validación y filtrado de datos provenientes del COES u otras fuentes, garantizando la integridad y confiabilidad de la información de entrada.
- Evaluación de ventanas de tiempo y configuración de entrada: Se recomienda probar distintos tamaños de ventana y configuraciones de input features, analizando su efecto en la estabilidad y capacidad de generalización del modelo a diferentes horizontes de predicción.
- Entrenamiento continuo y actualización periódica del modelo: Para mantener la precisión predictiva en el tiempo, se sugiere implementar esquemas de actualización periódica del modelo con nuevos datos, o adoptar metodologías de aprendizaje continuo (online learning) que permitan adaptarse dinámicamente a cambios en los patrones de demanda.
- Aplicación en usuarios libres sin y con BESS: El modelo puede ser implementado tanto en usuarios con sistemas de almacenamiento de

energía (BESS), mediante integración con sistemas EMS para automatizar estrategias de load shifting, como en usuarios sin almacenamiento, donde puede servir para programar reducciones de consumo y evitar penalidades.

- Desarrollo de plataformas visuales para toma de decisiones: Se recomienda implementar dashboards o interfaces gráficas intuitivas que permitan a los usuarios libres interpretar los resultados del modelo y tomar decisiones estratégicas de forma eficiente y en tiempo real.
- Evaluación del modelo ante eventos atípicos: Se sugiere realizar estudios adicionales sobre el comportamiento del modelo frente a eventos específicos como feriados, crisis energéticas, fluctuaciones económicas o cambios regulatorios, a fin de fortalecer su resiliencia en contextos no ordinarios.
- Consideración para políticas regulatorias: Se recomienda que organismos reguladores del sector energético consideren herramientas predictivas como la desarrollada en esta investigación para apoyar la planificación operativa, diseñar mecanismos de respuesta a la demanda y promover un uso más eficiente de la energía.
- Investigación de nuevas metodologías emergentes: Finalmente, se propone explorar nuevas técnicas de predicción, como redes generativas adversarias (GANs), modelos tipo transformer o redes neuronales informadas por física (Physics-Informed Neural Networks), con el fin de evaluar su aplicabilidad en el contexto eléctrico nacional.

Referencias Bibliográficas

- Apanian, T. (2018). *A Five-Minute Guide to Artificial Intelligence*. Retrieved from <https://medium.com/appanian/a-five-minute-guide-to-artificial-intelligence-c4262be85fd3>
- Chua, K. H., Bong, H. L., Lim, Y. S., Wong, J., & Wang, L. (2020). *The State-of-the-Arts of Peak Shaving Technologies: A Review*. Retrieved from <https://ieeexplore.ieee.org/abstract/document/9275637>
- COES. (2024). *Comité de Operación Económica del Sistema Interconectado Nacional*. Retrieved from <https://www.coes.org.pe/>
- Curo Martinez, J. (2022). *Pronóstico de la demanda eléctrica diaria del SEIN para reducir la demanda coincidente de un usuario libre del sector industrial*. Retrieved from <http://hdl.handle.net/20.500.12894/8492>
- Gallo Cruz. (2020). *Análisis predictivo para minería de datos y proyección a corto plazo de la demanda de potencia en el sistema eléctrico ecuatoriano*.
- Kluyver, T. R.-K. (2016). *Jupyter Notebooks*. Retrieved from A publishing format for reproducible computational workflows.
- Leva Apaza, A. (2023). *Diseño de un modelo computacional mediante redes neuronales artificiales para predecir consumos de energía eléctrica, Perú 2022*. Retrieved from <https://hdl.handle.net/20.500.12952/8171>

- Maza Chauca, J. L., & Pérez Félix, M. Á. (2022). *Almacenamiento de energía eléctrica en la central fotovoltaica de Moquegua y su inyección en horas de máxima demanda del SEIN*. Retrieved from <https://hdl.handle.net/20.500.14278/3852>
- OSINERGMIN. (2025). *Contratos de Usuarios Libres*. Retrieved from <https://www.osinergmin.gob.pe/empresas/electricidad/generacion/contratos-de-usuarios-libres>
- Pariona Curi, M. A. (2022). *Peak shaving y su influencia en la optimización de los costos de energía en un cliente libre del sector industrial de manufactura*.
- Sepp, H., & Jürgen, S. (1997). *Long Short-Term Memory*. Retrieved from Neural Comput: <https://doi.org/10.1162/neco.1997.9.8.1735>
- Taylor, S., & Letham, B. (2018). *Forecasting at Scale*. Retrieved from The American Statistician: <https://doi.org/10.1080/00031305.2017.1380080>

Anexos

Anexo A: Matriz de consistencia	1
Anexo B: Código R del modelo Sarima	2
Anexo C: Código Python del modelo Prophet	8
Anexo D: Código Python del modelo LSTM	12

Anexo A: Matriz de consistencia

	Problema de Investigación	Objetivos	Hipótesis	Variables	Preguntas de Investigación	Metodología
General	Los usuarios libres (UL) del SEIN enfrentan costos elevados de facturación debido a su consumo durante la Máxima Demanda Coincidente (MDC), sin contar con herramientas predictivas que les permitan anticipar los momentos en que ocurre la MDC y optimizar su consumo mediante estrategias como Peak Shaving o Load Shifting.	Desarrollar una herramienta predictiva basada en la prognosis de la generación de potencia del SEIN para identificar los momentos de la MDC y facilitar a los usuarios libres la optimización de sus costos de facturación.	La prognosis precisa de la generación de potencia del SEIN permite predecir los momentos más probables de la MDC, lo que facilita a los usuarios libres implementar estrategias de optimización del consumo y reducir sus costos de facturación.	V. Independiente: Prognosis de la generación de potencia del SEIN	1. ¿Qué patrones históricos de la generación diaria del SEIN son útiles para predecir los momentos de la MDC? 2. ¿Qué modelo predictivo es más efectivo para anticipar los momentos probables de la MDC? 3. ¿Cómo pueden los usuarios libres utilizar las predicciones de la MDC para implementar estrategias de optimización del consumo?	Enfoque: Cuantitativo. Diseño: No experimental, longitudinal y retrospectivo. Población: Datos históricos y actuales de generación del SEIN (mediciones en intervalos de 15 minutos durante todo el día). Muestra: Datos históricos de generación de al menos cinco años completo para capturar patrones estacionales. Instrumentos: Software para manejo y modelado de datos (Python, Jupyter Notebook y R). Técnicas de análisis: Evaluación de modelos predictivos con métricas como MAE, RMSE y precisión predictiva para determinar la fiabilidad del modelo. Validación cruzada con datos históricos para garantizar robustez del modelo. Resultados esperados: Herramienta predictiva que identifique los momentos más probables de la MDC y recomendaciones para la optimización del consumo de los usuarios libres.
Específico	1. No se dispone de una herramienta predictiva confiable para anticipar los momentos de la MDC. 2. La falta de información sobre los patrones de generación diaria dificulta la implementación de estrategias de optimización del consumo por parte de los UL. 3. Los altos costos asociados a potencia y peajes de transmisión afectan la competitividad económica de los usuarios libres.	1. Diseñar y evaluar modelos predictivos de prognosis de generación diaria para determinar los momentos probables de la MDC. 2. Analizar los patrones históricos de generación del SEIN para identificar tendencias que permitan anticipar los momentos de la MDC. 3. Validar la precisión de los modelos predictivos mediante datos históricos y actuales del SEIN. 4. Proponer recomendaciones prácticas para la implementación de estrategias de optimización de consumo por parte de los usuarios libres.	1. Los patrones históricos de generación contienen información relevante para predecir los momentos de la MDC. 2. Un modelo predictivo adecuado puede identificar los momentos más probables de la MDC con alta precisión. 3. La predicción precisa de la MDC permite a los usuarios libres aplicar estrategias de optimización, disminuyendo su exposición a altos costos de facturación por potencia y peajes de transmisión.	V. Dependiente: Costos de facturación de los usuarios libres.		

Anexo B: Código R del modelo Sarima

```
#
=====

# Modelo SARIMA para la Predicción de la Generación Eléctrica en el SEIN

# Autor: Ing. Erick Del Piero Y Daniel Gonzales Requejo

# Descripción: Este script ajusta un modelo SARIMA a la serie temporal de

# generación eléctrica, evalúa su desempeño y genera predicciones.

#
=====

# Carga de librerías necesarias

library(forecast) # Para modelado de series temporales y predicción

library(ggplot2) # Para visualización de datos

library(readr) # Para la lectura de archivos CSV

#
=====

# 1. Carga de Datos

#
=====

# Cargar los datos desde un archivo CSV

data
read_csv("/home/erickdelpiero/Documents/Tesis/2.Modelado/Analisis/Generación/Genera
cion.csv")

# Creación de la serie temporal
```

```

# - La serie representa la potencia generada en el SEIN

# - Frecuencia: 96 (muestras por día, equivalente a registros cada 15 minutos)

# - Inicio: (1,1) (definido arbitrariamente)

serie_ts <- ts(data$Potencia, frequency = 96, start = c(1, 1))

#
=====

# 2. División de Datos en Conjunto de Entrenamiento y Prueba

#
=====

# Determinar el tamaño de la serie temporal

n <- length(serie_ts)

# División de la serie:

# - Conjunto de entrenamiento: Hasta el último día completo antes del período de
prueba

# - Conjunto de prueba: Desde el primer día del período de evaluación en adelante

train <- window(serie_ts, end = c(2191, 96))

test <- window(serie_ts, start = c(2192, 1))

#
=====

# 3. Ajuste del Modelo SARIMA

#
=====

# Ajuste automático de un modelo SARIMA a los datos de entrenamiento

```

```

modelo_train <- auto.arima(train, seasonal = TRUE)

# Resumen del modelo ajustado

summary(modelo_train)

#
=====

# 4. Evaluación del Modelo

#
=====

# Generación de predicciones para el conjunto de prueba

predicciones <- forecast(modelo_train, h = length(test))

# Cálculo de métricas de precisión

accuracy(predicciones, test)

# Cálculo del MAPE (Error Porcentual Medio Absoluto)

mape <- mean(abs((test - predicciones$mean) / test)) * 100

print(paste("MAPE:", round(mape, 2), "%"))

#
=====

# 5. Almacenamiento del Modelo Ajustado

#
=====

# Guardar el modelo en formato .RData

save(modelo_train,
      file
      =
      "/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Modelo/modelo_train.RData"
)

```

```

# Guardar el modelo en formato .rds

saveRDS(modelo_train, file =
"/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Modelo/modelo_train.rds")

#
=====

# 6. Generación de Predicciones a Largo Plazo

#
=====

# Predicción para los próximos 30 días (96 registros por día)

predicciones <- forecast(modelo_train, h = 30 * 96)

# Visualización de las predicciones

plot(predicciones)

# Exportar predicciones a un archivo CSV

write.csv(predicciones,
"/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Predicciones_SARIMA.csv")

#
=====

# 7. Análisis de Residuos del Modelo

#
=====

# Obtención de los residuos del modelo ajustado

residuos <- residuals(modelo_train)

# 7.1. Gráfico de Residuos en el Tiempo

```



```

    ggplot(data = data.frame(Time = time(residuos), Residuals = residuos), aes(x =
Time, y = Residuals)) +

    geom_line(color = "blue") +

    labs(title = "Residuos del Modelo SARIMA", x = "Tiempo", y = "Residuos") +

    theme_minimal()

# 7.2. Función de Autocorrelación (ACF) de los Residuos

ggAcf(residuos, main = "Autocorrelación (ACF) de los Residuos")

# 7.3. Función de Autocorrelación Parcial (PACF) de los Residuos

ggPacf(residuos, main = "Autocorrelación Parcial (PACF) de los Residuos")

# 7.4. Prueba de Ljung-Box para verificar si los residuos son ruido blanco

# - Se usa un lag de 96 para considerar un ciclo diario completo

Box.test(residuos, lag = 96, type = "Ljung-Box")

#
=====

# 8. Almacenamiento de Gráficos de Diagnóstico

#
=====

# Guardar gráfico de residuos en el tiempo

ggsave("/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Performance/
2017/residuos_tiempo.png", width = 10, height = 6)

# Guardar gráfico ACF de residuos

ggsave("/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Performance/
2017/acf_residuos.png",

    plot = ggAcf(residuos), width = 10, height = 6)

```

```

# Guardar gráfico PACF de residuos

ggsave("/home/erickdelpiero/Documents/Tesis/2.Modelado/SARIMA/Performance/
2017/pacf_residuos.png",

plot = ggPacf(residuos), width = 10, height = 6)

#
=====
=====

# Fin del Script

#
=====
=====

```

Anexo C: Código Python del modelo Prophet

```
#
=====

# Modelo Prophet para la Predicción de la Generación Eléctrica en el SEIN
# Autor: Ing. Erick Del Piero Y Daniel Gonzales Requejo
# Descripción: Este script ajusta un modelo Prophet a la serie temporal de generación
# eléctrica, evalúa su desempeño y genera predicciones.

#
=====

# Paso 1: Importar las librerías necesarias
import pandas as pd
from prophet import Prophet
from sklearn.metrics import mean_absolute_error, mean_squared_error
import matplotlib.pyplot as plt
import numpy as np
import joblib

#
=====

# 1. Carga de Datos

#
=====

# Definir la ruta del archivo CSV con los datos de generación eléctrica
data_path =
'/home/erickdelpiero/Documents/Tesis/2.Modelado/Análisis/Generación/Generacion.csv'

# Leer los datos en un DataFrame
data = pd.read_csv(data_path)

# Mostrar las primeras filas del DataFrame para verificar la carga correcta
data.head()

#
=====
```

2. Preprocesamiento de Datos

```
#
=====

# Convertir la columna de fecha y hora a formato datetime
data['FechaHora'] = pd.to_datetime(data['FechaHora'])

# Renombrar columnas según los requerimientos de Prophet:
# - 'FechaHora' se convierte en 'ds' (date-time stamp)
# - 'Potencia' se convierte en 'y' (variable a predecir)
data = data.rename(columns={'FechaHora': 'ds', 'Potencia': 'y'})

# Mantener únicamente las columnas necesarias para Prophet
data = data[['ds', 'y']]

# Verificar la estructura del DataFrame después del preprocesamiento
data.head()

data.tail()

#
=====
```

3. División de Datos en Entrenamiento y Prueba

```
#
=====

# Definir los conjuntos de entrenamiento y prueba:
# - Entrenamiento: Datos entre 2017 y el inicio de 2023
# - Prueba: Datos desde el inicio de 2023 hasta julio de 2024
train = data[(data['ds'] >= '2017-01-01') & (data['ds'] <= '2023-01-01 04:45:00')]
test = data[(data['ds'] >= '2023-01-01 05:00:00') & (data['ds'] <= '2024-07-01 04:45:00')]

# Mostrar las dimensiones de los conjuntos
print("Tamaño del conjunto de entrenamiento:", train.shape)
print("Tamaño del conjunto de prueba:", test.shape)

#
=====
```

4. Creación y Entrenamiento del Modelo Prophet

```

#
=====
=====

# Inicializar el modelo Prophet con configuración de estacionalidad
model = Prophet(
    seasonality_mode='additive', # Se puede cambiar a 'multiplicative' si es necesario
    yearly_seasonality=True,     # Considera la estacionalidad anual
    daily_seasonality=True,      # Considera la estacionalidad diaria
    weekly_seasonality=True      # Considera la estacionalidad semanal
)

# Entrenar el modelo con los datos de entrenamiento
model.fit(train)

#
=====
=====

# 5. Guardado del Modelo Entrenado

#
=====
=====

# Especificar la ruta donde se guardará el modelo entrenado
model_path =
'/home/erickdelpiero/Documents/Tesis/2.Modelado/Prophet/prophet_model.pkl'

# Guardar el modelo en un archivo .pkl para uso futuro
joblib.dump(model, model_path)

print(f"Modelo guardado exitosamente en: {model_path}")

#
=====
=====

# 6. Generación de Predicciones

#
=====
=====

# Crear un DataFrame con las fechas futuras a predecir
# - Se genera una proyección para el mismo período que el conjunto de prueba
future = model.make_future_dataframe(periods=len(test), freq='15min')

```

```
#
=====
=====

# Fin del Script

#
=====
=====
```

Anexo D: Código Python del modelo LSTM

```
#
=====

# Modelo LSTM para la Predicción de la Generación Eléctrica en el SEIN

# Autor: Ing. Erick Del Piero Y Daniel Gonzales Requejo

# Descripción: Este script entrena un modelo LSTM para la predicción de

# generación eléctrica en el SEIN utilizando múltiples variables de entrada.

#
=====

# Paso 1: Importar las librerías necesarias

import pandas as pd

import numpy as np

from sklearn.preprocessing import MinMaxScaler

import matplotlib.pyplot as plt

from keras.models import Sequential, load_model

from keras.layers import LSTM, Dense, Dropout

from sklearn.model_selection import train_test_split

from keras.callbacks import EarlyStopping

import os

#
=====

# 1. Carga y Preprocesamiento de Datos
```

```

#
=====

# Definir rutas de los archivos CSV con los datos de entrada

data_paths = {

    'Generacion':
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Variables/Generacion.csv",

    'Ejecutado':
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Variables/Ejecutado.csv",

    'Programado':
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Variables/Programado.csv",

    'Reprogramado':
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Variables/Reprogramado.csv",

    'Intercambio':
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Variables/Ecuador.csv"

}

# Cargar y procesar los datos en un diccionario

data_frames = {}

for key, path in data_paths.items():

    df = pd.read_csv(path)

    df['FechaHora'] = pd.to_datetime(df['FechaHora']) # Convertir a datetime

    df.set_index('FechaHora', inplace=True) # Establecer la columna de fecha como
índice

    data_frames[key] = df

# Normalización de datos (escala 0-1) utilizando MinMaxScaler

```



```

scalers = {}

for key in data_frames.keys():

    scaler = MinMaxScaler(feature_range=(0, 1))

    column = list(data_frames[key].columns)[0] # Obtener el nombre de la columna
de datos

    data_frames[key][column + '_scaled'] =
scaler.fit_transform(data_frames[key][[column]])

    scalers[key] = scaler # Guardar el scaler para futura desnormalización

#
=====

# 2. Construcción del Conjunto de Datos

#
=====

# Definir parámetros de la ventana de entrada y la predicción

window_size = 2688 # 96 pasos por día * 28 días (~4 semanas)

prediction_steps = 672 # 672 pasos (1 semana)

# Unir las variables en un único DataFrame alineado temporalmente

data_combined = pd.DataFrame({

    'Generacion': data_frames['Generacion']['Potencia_scaled'],

    'Ejecutado': data_frames['Ejecutado']['COES_scaled'],

    'Programado': data_frames['Programado']['PROGRAMADO_scaled'],

    'Reprogramado': data_frames['Reprogramado']['REPROGRAMADO_scaled'],

    'Intercambio': data_frames['Intercambio']['ECUADOR_scaled']

```

```

    })

    # Eliminar valores nulos (si los hay)

    data_combined.dropna(inplace=True)

    # Función para crear secuencias de datos para el modelo LSTM

    def create_sequences(data, target_column, window_size, prediction_steps):

        sequences, targets = [], []

        for i in range(len(data) - window_size - prediction_steps):

            sequences.append(data.iloc[i:i + window_size].values)

            targets.append(data.iloc[i + window_size:i + window_size +
prediction_steps][target_column].values)

        return np.array(sequences), np.array(targets)

    # Generar las secuencias de entrada y salida

    X, y = create_sequences(data_combined, target_column='Generacion',
window_size=window_size, prediction_steps=prediction_steps)

    # División en conjuntos de entrenamiento y prueba (85% train, 15% test)

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15,
random_state=42, shuffle=False)

    #
=====
=====

    # 3. Definición y Entrenamiento del Modelo LSTM

    #
=====
=====

    # Construcción del modelo secuencial LSTM

    model = Sequential([

```

```

        LSTM(50, return_sequences=True, input_shape=(X_train.shape[1],
X_train.shape[2])),

        Dropout(0.2),

        LSTM(50, return_sequences=False),

        Dropout(0.2),

        Dense(672) # Capa de salida con el número de pasos de predicción

    ])

# Compilar el modelo con función de pérdida MSE y optimizador Adam

model.compile(optimizer='adam', loss='mean_squared_error')

# Configurar el Early Stopping para evitar sobreajuste

early_stopping = EarlyStopping(monitor='val_loss', patience=10,
restore_best_weights=True)

# Entrenar el modelo

history = model.fit(

    X_train, y_train,

    validation_data=(X_test, y_test),

    epochs=50,

    batch_size=32,

    callbacks=[early_stopping],

    verbose=1

)

#
=====
=====

```

4. Evaluación del Modelo y Guardado

```
#
=====

# Evaluar la pérdida en los conjuntos de entrenamiento y prueba

train_loss = model.evaluate(X_train, y_train, verbose=1)

test_loss = model.evaluate(X_test, y_test, verbose=1)

print(f"Train Loss: {train_loss}, Test Loss: {test_loss}")

# Guardar el modelo entrenado

model_save_path =
"/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Modelos/Modelo.h5"

model.save(model_save_path)

print(f"Modelo guardado en: {model_save_path}")

#
=====

# 5. Visualización del Entrenamiento

#
=====

# Graficar la evolución de la pérdida durante el entrenamiento

plt.figure(figsize=(10, 6))

plt.plot(history.history['loss'], label='Pérdida de entrenamiento')

plt.plot(history.history['val_loss'], label='Pérdida de validación')

plt.legend()

plt.xlabel('Épocas')
```

```

plt.ylabel('Pérdida')

plt.title('Evolución de la pérdida durante el entrenamiento')

plt.grid()

# Guardar la gráfica de pérdida

plot_save_path = '/home/erickdelpiero/Documents/Tesis/7.LSTMv2/Graficas/V1'

os.makedirs(plot_save_path, exist_ok=True)

plt.savefig(os.path.join(plot_save_path, 'loss.png'))

plt.show()

#
=====

# 6. Generación de Predicciones

#
=====

# Definir un intervalo de prueba para predicción

interval_start = 0

interval_end = 672 # 672 timestamps (1 semana)

X_subset = X_test[interval_start:interval_end]

y_subset = y_test[interval_start:interval_end]

# Realizar predicciones con el modelo entrenado

predictions = model.predict(X_subset)

# Desnormalizar las predicciones y los valores reales

scaler_generacion = scalers['Generacion'] # Usar el scaler original

```

```

predictions_descaled = scaler_generacion.inverse_transform(predictions)

1)) y_subset_descaled = scaler_generacion.inverse_transform(y_subset.reshape(-1,

# Graficar Predicción vs Realidad

plt.figure(figsize=(12, 8))

plt.plot(y_subset_descaled, label='Valor Real', color='blue')

plt.plot(predictions_descaled, label='Predicción', color='orange')

plt.legend()

plt.xlabel('Timestamps')

plt.ylabel('Generación (MW)')

plt.title('Comparación: Predicción vs Realidad')

plt.grid()

# Guardar la gráfica de predicción

plt.savefig(os.path.join(plot_save_path, 'prediction.png'))

plt.show()

#
=====
=====

# Fin del Script

#
=====
=====

```