



UNIVERSIDAD NACIONAL DE CHIMBORAZO

FACULTAD DE INGENIERÍA

CARRERA DE TELECOMUNICACIONES

Desarrollo de un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba, mediante la aplicación de algoritmos de predicción, para la mejora de la gestión energética.

Trabajo de Titulación para optar por el título de

INGENIERO(A) EN TELECOMUNICACIONES

Autor(a):

Victoria Estefanía Parra Rocha

Tutor:

PhD. Ciro Diego Radicelli García

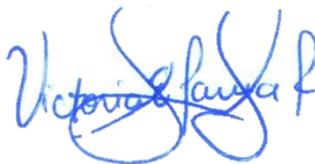
Riobamba, Ecuador. 2025

DERECHOS DE AUTORÍA

Yo, Victoria Estefanía Parra Rocha, con cédula de ciudadanía 0604537779, autora del trabajo de investigación titulado: Desarrollo de un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba, mediante la aplicación de algoritmos de predicción, para la mejora de la gestión energética, certifico que la producción, ideas, opiniones, criterios, contenidos y conclusiones expuestas son de mí exclusiva responsabilidad.

Asimismo, cedo a la Universidad Nacional de Chimborazo, en forma no exclusiva, los derechos para su uso, comunicación pública, distribución, divulgación y/o reproducción total o parcial, por medio físico o digital; en esta cesión se entiende que el cesionario no podrá obtener beneficios económicos. La posible reclamación de terceros respecto de los derechos de autor (a) de la obra referida, será de mi entera responsabilidad; librando a la Universidad Nacional de Chimborazo de posibles obligaciones.

En Riobamba, el 24 de febrero de 2025



Victoria Estefanía Parra Rocha

C.I: 0604537779



ACTA FAVORABLE – INFORME FINAL DEL TRABAJO DE INVESTIGACIÓN

En la ciudad de Riobamba, a los 25 días del mes de FEBRERO de 2025, luego de haber revisado el Informe Final del Trabajo de Investigación presentado por la estudiante **PARRA ROCHA VICTORIA ESTEFANÍA** con CC: **060453777-9**, de la carrera de **INGENIERÍA EN TELECOMUNICACIONES** y dando cumplimiento a los criterios metodológicos exigidos, se emite el **ACTA FAVORABLE DEL INFORME FINAL DEL TRABAJO DE INVESTIGACIÓN** titulado **“DESARROLLO DE UN SISTEMA PREDICTIVO DE CONSUMO DE ENERGÍA ELÉCTRICA EN LA CIUDAD DE RIOBAMBA, MEDIANTE LA APLICACIÓN DE ALGORITMOS DE PREDICCIÓN, PARA LA MEJORA DE LA GESTIÓN ENERGÉTICA”**, por lo tanto se autoriza la presentación del mismo para los trámites pertinentes.



Firmado electrónicamente por:
**CIRO DIEGO
RADICELLI GARCIA**

PhD. **Ciro Diego Radicelli García**
TUTOR

CERTIFICADO DE LOS MIEMBROS DEL TRIBUNAL

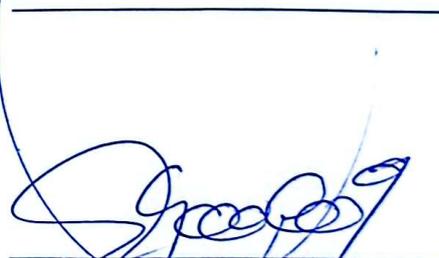
Quienes suscribimos, catedráticos designados Miembros del Tribunal de Grado del trabajo de investigación **Desarrollo de un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba, mediante la aplicación de algoritmos de predicción, para la mejora de la gestión energética por Victoria Estefanía Parra Rocha**, con cédula de identidad número **060453777-9**, bajo la tutoría de **PhD. Ciro Diego Radicelli Garcia**; certificamos que recomendamos la **APROBACIÓN** de este con fines de titulación. Previamente haber revisado y evaluado el trabajo de investigación y cumplida la sustentación por parte de su autor; no teniendo más nada que observar.

De conformidad a la normativa aplicable firmamos, en Riobamba, a la fecha de su presentación el 12 de mayo de 2025.

PhD. Marlon Danilo Basantes Valverde
PRESIDENTE DEL TRIBUNAL DE GRADO



Msc. Edgar Giovanni Cuzco Silva
MIEMBRO DEL TRIBUNAL DE GRADO



PhD. Carlos Ramiro Peñafiel Ojeda
MIEMBRO DEL TRIBUNAL DE GRADO





CERTIFICACIÓN

Que, **PARRA ROCHA VICTORIA ESTEFANÍA** con CC: **060453777-9**, estudiante de la Carrera de **TELECOMUNICACIONES**, Facultad de **INGENIERÍA**; ha trabajado bajo mi tutoría el trabajo de investigación titulado **“DESARROLLO DE UN SISTEMA PREDICTIVO DE CONSUMO DE ENERGÍA ELÉCTRICA EN LA CIUDAD DE RIOBAMBA, MEDIANTE LA APLICACIÓN DE ALGORITMOS DE PREDICCIÓN, PARA LA MEJORA DE LA GESTIÓN ENERGÉTICA”**, cumple con el **5%** de acuerdo al reporte del sistema Anti plagio **COMPILATIO**, porcentaje aceptado de acuerdo a la reglamentación institucional, por consiguiente autorizo continuar con el proceso.

Riobamba, 23 de abril de 2025.



Firmado electrónicamente por:
CIRO DIEGO
RADICELLI GARCIA

PhD. Ciro Diego Radicelli García
TUTOR

DEDICATORIA

Este trabajo está dedicado a mi yo del pasado, a mi yo del presente y a mi yo del futuro; quiero recordarnos que no hay adversidad que no se supere; que ningún obstáculo es tan grande como para no ser capaz de cruzarlo y que los que no cruzamos no nos llevaban, ni nos llevarán al camino correcto; que somos capaces de lograr lo que nos proponemos y que nuestro límite no es el cielo, va más allá de eso y nuestro propósito es encontrarlo a través de causas que nos hagan verdaderamente felices.

Victoria Estefanía Parra Rocha

AGRADECIMIENTO

Durante el tiempo de ejecución de este proyecto, varias personas han sido fundamentales, no solo por el interés que han puesto en esto como un logro propio, si no también por la paciencia, motivación y apoyo que le han dado a su autora.

Agradezco a mi mami, Luisa Rocha, que siempre demostró su interés y no permitió que dejara mis sueños de lado, que nunca me niega su apoyo y siempre premia mis logros, por muy pequeños que sean.

Agradezco a mi tía, Ene Rocha; quien nunca me dejó sola, siempre se preocupó de mi bienestar y junto a mi perrito, Max, que sin importar lo largos que parecieran los días, dormía a lado de mi trabajo, contemplando que continúe como de costumbre en el lugar de siempre.

Agradezco a mi pa, Juan; a mi ñaña, Andre-ita, a mis abuelitos, Luz, Fausto y Lupe, porque todos los días ven por mi bienestar, no permiten nunca que me aleje del camino que elegí, y quienes incansablemente me han aconsejado y aconsejarán.

Gracias a mi tutor, Doctor Ciro Radicelli, de quien he sentido una absoluta confianza desde el primer día, y ha aportado ilimitadamente con su conocimiento a mi camino y al de muchos como yo, quiero expresar que su validación me llena de orgullo.

Por último y no menos importante, el favorito; gracias Juan Francisco por no dejar que me hunda en esto, gracias por la compañía y los recordatorios durante mi trabajo, gracias por alegrarme en los días más oscuros.

Quiero agregar un agradecimiento a todos quienes me han acompañado durante mi tiempo de formación, quienes han aportado a mi vida de una u otra forma, con una mención especial a mis compañeros y profesores, nunca nadie llega a sus metas solo, gracias por las personas que han respaldado mis pasos.

Gracias infinitas,

Victoria Estefanía Parra Rocha

ÍNDICE GENERAL

DERECHOS DE AUTORÍA

DICTAMEN FAVORABLE DEL PROFESOR TUTOR

CERTIFICADO DE LOS MIEMBROS DEL TRIBUNAL

CERTIFICADO ANTIPLAGIO

DEDICATORIA

AGRADECIMIENTO

RESUMEN

ABSTRACT

CAPÍTULO I	19
1.1. INTRODUCCIÓN	19
1.2. Antecedentes	19
1.3. Formulación del Problema.....	20
1.4. Objetivos	20
1.4.1 General.....	20
1.4.2 Específicos.....	21
CAPÍTULO II. MARCO TEÓRICO.....	22
2.1. Estado del Arte	22
2.2 Marco Teórico	24
2.2.1 Energía Eléctrica.....	24
2.2.2 Gestión de la Energía Eléctrica	25
2.2.3 Empresa Eléctrica Riobamba S.A.	25
2.2.4 Machine Learning.....	26
2.2.5 Algoritmos de Machine Learning.....	27
2.2.6 Modelos de Machine Learning.....	28

2.2.7	Series Temporales.....	29
2.2.8	Redes Neuronales Artificiales	30
2.2.9	Redes Neuronales Recurrentes	31
2.2.10	ARIMA y ARIMAX.....	32
2.2.11	Modelos Híbridos	34
2.2.12	Variables Exógenas	34
2.2.13	Métricas de Error	34
CAPÍTULO III. METODOLOGÍA.....		36
3.1.	Tipo de Investigación	36
3.2.	Métodos de Investigación.....	36
3.3	Enfoque de la Investigación	36
3.4.	Proceso metodológico	37
3.5.	Población y Muestra	37
3.5.1	Población	37
3.5.2	Muestra	38
3.6	Técnicas e Instrumentos	38
3.6.1	Técnicas de la Investigación.....	38
3.6.2	Instrumentos de la Investigación	39
3.7.	Operacionalización de variables	39
3.8	Pre procesamiento de Datos	41
3.8.1	Limpieza de Datos.....	41
3.8.2	Reducción de Dimensionalidad	45
3.8.3	Integración y Fusión de Datos	46
3.8.4	División del Conjunto de Datos	47

3.9. Estacionalidad de los Datos.....	47
3.10 Relación entre el Consumo y las variables exógenas (días de la semana, feriados y temperatura).....	49
3.10.1 Relación entre el Consumo y la Temperatura	49
3.10.2 Relación entre el Consumo y los días de la semana	50
3.10.3 Relación entre el Consumo y la condición de feriado	50
3.11. Entrenamiento del Modelo	50
3.11.1 Desarrollo del modelo SARIMAX	51
3.11.2 Desarrollo del modelo híbrido (SARIMAX con Redes Neuronales)	56
3.11.3 Desarrollo del modelo LSTM (Long Short-Term Memory)	59
3.12 Validez del modelo mejor ajustado	62
CAPÍTULO IV. RESULTADOS Y DISCUSIÓN.....	64
4.1. Estacionalidad del Consumo Eléctrico.....	64
4.2 Análisis de Variables Exógenas	64
4.3 Comparación de Modelos.....	65
4.4 Evaluación del Modelo LSTM	65
4.5 Validación del modelo.....	68
4.6 Implicaciones y Aplicabilidad.....	68
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.	70
5.1 Conclusiones.....	70
5.2 . Recomendaciones	72
BIBLIOGRAFÍA	74
ANEXOS	77
Anexo 1. Carta de Compromiso	77
Anexo 2. Registro de Desarrollo del modelo SARIMAX	78

Anexo 3. Registro de Desarrollo del modelo SARIMAX+RNN	81
Anexo 4. Registro de Desarrollo del modelo LSTM.....	83

ÍNDICE DE TABLAS

Tabla 1.- Características claves de la Energía Eléctrica [13]	25
Tabla 2.- Modelos de Machine Learning. [18].....	29
Tabla 3.- Operacionalización de Variables. Variables dependiente e independiente.....	40
Tabla 4.- Operacionalización de Variables. Variables exógenas.	41
Tabla 5.- Outliers modificados en la base de datos.	44
Tabla 6.- Formato de la base de datos proporcionada por la fuente.....	46
Tabla 7.- Parámetros para los modelos SARIMAX y SARIMAX+RNN.	54
Tabla 8.- MAE, MAPE y RMSE para modelo SARIMAX.....	55
Tabla 9.- MAE, MAPE y RMSE para el modelo de SARIMAX + RNN.	59
Tabla 10.- MAPE, MAE y RMSE para el modelo LSTM.	62
Tabla 11.- Interpretación de los errores.....	63
Tabla 12.-Comparación de errores MAE, MAPE y RMSE para los modelos planteados.	65
Tabla 13.- Resultados de la predicción realizada con el modelo LSTM.....	67

ÍNDICE DE FIGURAS

Figura 1.- Mapa de la infraestructura eléctrica de la ciudad de Riobamba. [16].....	26
Figura 2.- Algoritmos de Machine Learning. [17]	27
Figura 3.- Fases de modelado de algoritmos. [18]	29
Figura 4.- Estructura jerárquica de un sistema basado en RNA. [21]	31
Figura 5.- Proceso de la Metodología.....	37
Figura 6.- Método del Codo para identificar el número adecuado de clústeres.	42
Figura 7.-Gráfica de Clústeres tomando en cuenta las variables exógenas.....	42
Figura 8.- Diagrama de caja o boxplot por clústeres.....	43
Figura 9.- Tabla de outliers generada por el método de Clustering.	44
Figura 10.- Consumo de 10 días aleatorios en el conjunto de datos.	48
Figura 11.- Consumo promedio diario de energía eléctrica en la ciudad de Riobamba en kWh	48
Figura 12.- Gráfica de dispersión de los datos de Consumo vs Temperatura.	49
Figura 13.- Media de Consumo de los días entre semana vs los fines de semana.	50
Figura 14.- Media de Consumo vs condición de feriado.....	50
Figura 15.- Gráfica PACF para determinar parámetros ARIMA.	52
Figura 16.- Gráfica ACF para determinar parámetros ARIMA.	53
Figura 17.- Predicción Consumo Real vs Consumo Predicho con el modelo SARIMAX (0,1,0(0,0,0).	55
Figura 18.- Arquitectura de la Red Neuronal Recurrente (RNN) utilizada en el modelo híbrido.....	57
Figura 19.- Comparación del consumo real vs predicción del modelo híbrido.....	59
Figura 20.- Arquitectura del modelo LSTM.....	60
Figura 21.- Comparación del Consumo Real vs la predicción utilizando LSTM.	62

Figura 22.- Comparación entre los consumos reales y la predicción utilizando el modelo seleccionado.....	66
Figura 23.- Comparación entre el consumo real y el consumo predicho por el modelo LSTM para el 21 de septiembre de 2023.	67

ÍNDICE DE ECUACIONES

(1)	33
(2)	58

RESUMEN

El presente estudio tiene como objetivo desarrollar un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba, mediante la aplicación de modelos de series temporales y redes neuronales, con el propósito de mejorar la gestión energética y asistir en la toma de decisiones con respecto a la crisis energética que atraviesa el país.

Para ello, se analizaron datos históricos de consumo eléctrico correspondientes al periodo 2021-2023, con una frecuencia de 15 minutos, lo que resultó en un total de 105,320 registros. Las variables consideradas incluyen el consumo eléctrico como variable dependiente, el modelo utilizado como variable independiente y tres variables exógenas: la temperatura, el día de la semana y las festividades o feriados.

En el desarrollo del modelo se evaluaron distintos enfoques predictivos, destacando el uso de SARIMAX, SARIMAX combinado con redes neuronales y modelos LSTM. Los resultados indicaron que el modelo LSTM proporcionó la mejor precisión en la predicción, basándose en métricas de error como el Error Absoluto Medio (MAE), el Error Porcentual Absoluto Medio (MAPE) y el Error Cuadrático Medio (RMSE). Sin embargo, se identificaron desafíos en la estimación de parámetros adecuados para los modelos ARIMAX y SARIMAX, así como en el manejo de grandes volúmenes de datos para redes neuronales.

Finalmente, el estudio proporciona recomendaciones para la mejora de la gestión del consumo energético en la ciudad, con base en los patrones identificados y las predicciones generadas, lo que contribuye a una mejor planificación del suministro eléctrico.

Palabras clave: SARIMAX, Redes Neuronales, LSTM, consumo eléctrico, gestión energética.

ABSTRACT

The following research aims to develop a predictive system for electric energy consumption in Riobamba by applying time series models and neural networks. The system aims to improve energy management and assist in Ecuador's energy crisis decision-making. Therefore, historical data on electric energy consumption from 2021–2023 was analyzed every 15 minutes, resulting in 105,320 records. The considered variables include electric consumption as the dependent variable, the model used as the independent variable, and three exogenous variables: temperature, type of day, and holidays. Different predictive approaches were evaluated during the model's development, highlighting the use of SARIMAX, SARIMAX combined with neural networks, and LSTM models. The results indicated that the LSTM model provided the highest prediction accuracy based on error metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). However, it was possible to identify challenges in estimating appropriate parameters for ARIMAX and SARIMAX models and handling large datasets for neural networks. Finally, the study provides recommendations for improving energy consumption management in the city based on the identified patterns and generated predictions, contributing to better electricity supply planning.

Key words: SARIMAX, Neural networks, LSTM, energy consumption, energy management.



Reviewed by:

Mgs. Jessica María Guaranga Lema
ENGLISH PROFESSOR
C.C. 0606012607

CAPÍTULO I

1.1. INTRODUCCIÓN

En la actualidad, la planificación y gestión de la oferta y demanda de energía eléctrica en Riobamba, así como en Ecuador en general, enfrenta desafíos significativos debido a la variabilidad en los patrones de consumo y el estiaje que atraviesa el país [1]. A fin de contrarrestar esta problemática, el gobierno ha propuesto medidas como la promoción de la inversión privada en la industria eléctrica y la contratación de generación extra [2]. Sin embargo, surge un problema adicional: la falta de herramientas predictivas precisas y adaptadas a las características específicas de la ciudad, resultando en un uso ineficiente de los recursos energéticos y posibles escaseces o excedentes de energía. Esta situación impacta la estabilidad del suministro eléctrico y provoca elevación de costos [3].

Ante estos desafíos, el gobierno propone medidas como la promoción de inversión privada y contratación de generación adicional para afrontar el período de estiaje 2024-2025 [2]. La contratación de energía privada requiere estimaciones precisas de la cantidad necesaria por período para presupuestar la compra. Además, la falta de un plan de gestión de energía en Ecuador plantea la necesidad de determinar si el país consume más energía de la que produce o si se pierde en su gestión.

El presente trabajo tiene como objetivo estudiar, adaptar e implementar algoritmos predictivos existentes a la realidad de la Empresa Eléctrica Riobamba S.A. de manera que se pueda predecir el consumo de la energía eléctrica en la ciudad de Riobamba y así mejorar la gestión de la energía y facilitar la adopción de nuevas medidas frente a la problemática que atraviesa el sector en la actualidad.

1.2. ANTECEDENTES

El 18 de octubre del año 2023, el Ministerio de Energía y Minas del Ecuador declaró en Emergencia al Sector Eléctrico en todo el territorio nacional, esto para precautelar que el suministro eléctrico se mantenga estable en todo el territorio con el objetivo de acelerar procesos de contratación de energía y optimización de recursos de las empresas eléctricas del país [1].

Es bien sabido que el país produce y contrata energía para abastecer el servicio eléctrico en todo el Ecuador, sin embargo debido al fenómeno del Niño, que ha provocado bajos caudales en los afluentes de la Amazonia, y un retraso de la época de lluvias en la zona occidental, la venta de energía por parte de Colombia tiene una limitación y un considerable incremento en su costo, la electricidad que Colombia ha estado vendiendo a Ecuador ha alcanzado precios de hasta \$0,60 por kilovatio hora [4] ; además de esto, el país ha registrado un crecimiento en la

demanda de este servicio; según datos del Operador Nacional de Electricidad (CENACE), la demanda ha crecido un 12% a octubre de 2023 frente al mismo período de 2022 [2]; razones por las que el Gobierno del Ecuador ha decretado la suspensión del servicio eléctrico en períodos determinados de tiempo en todo el territorio nacional con el objetivo de salvaguardar la estabilidad del suministro eléctrico.

1.3. FORMULACIÓN DEL PROBLEMA

El gobierno del Ecuador ha pensado y manifestado dos medidas principales para contrarrestar los apagones que sufre el país actualmente debido al período de estiaje en el que se encuentra, estas son: priorizar la promoción de la inversión privada en la cadena de la industria eléctrica e iniciar los procesos de contratación de generación adicional para abastecer la demanda eléctrica del país, principalmente para afrontar el período de estiaje 2024-2025 [2].

La contratación de energía al sector privado se entiende como la compra de energía por parte de las empresas eléctricas a empresas privadas que generan y almacenan su propia energía, esto conlleva a que las empresas públicas deberán contar con una estimación de cuanta energía eléctrica necesitarán por período de manera que se pueda presupuestar la compra de la misma dentro de la asignación de presupuesto a estas empresas.

El estudio de la demanda eléctrica contempla una proyección vigente para el período 2018-2027 sobre consumo de energía en el país, sin embargo, este se lo ha realizado con modelos econométricos [5], por lo que es importante emplear modelos y algoritmos avanzados dirigidos a áreas energéticas que tengan en cuenta las peculiaridades locales, como patrones estacionales específicos, temperatura, eventos particulares que afectan el consumo y cambios en la actividad económica. Para la implementación de medidas como las mencionadas, es necesario un pronóstico confiable de consumo de manera que se pueda realizar presupuestos a través de este.

Por lo tanto, el problema principal radica en brindar opciones para combatir esta emergencia mediante el desarrollo e implementación de algoritmos predictivos avanzados que puedan proporcionar pronósticos precisos y oportunos del consumo de energía eléctrica en la ciudad de Riobamba.

1.4. OBJETIVOS

1.4.1 General

- Desarrollar un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba, mediante la aplicación de algoritmos de predicción, para la optimización de la gestión energética

1.4.2 Específicos

- Analizar el histórico de datos del consumo de energía eléctrica correspondiente a los años 2021, 2022, 2023 en la ciudad de Riobamba, identificando variables relevantes que influyen en el consumo de energía.
- Estudiar y seleccionar los algoritmos predictivos a ser utilizados en el desarrollo del sistema predictivo utilizando los datos históricos obtenidos.
- Evaluar la precisión y fiabilidad del sistema predictivo comparándolo con datos reales de consumo.
- Proponer recomendaciones prácticas en pro de mejorar la gestión energética en la ciudad de Riobamba.

CAPÍTULO II. MARCO TEÓRICO.

2.1. ESTADO DEL ARTE

Según el informe emitido por la Administración de Información Energética (EIA) de EE.UU. [6], hasta 2040 puede producirse un crecimiento del 28% en la demanda mundial de energía. Es por eso que Fayaz y Kim [7], en su artículo titulado "A Prediction Methodology of Energy Consumption Based on Deep Extreme Learning Machine and Comparative Analysis in Residential Buildings", indican que es importante controlar de manera eficaz el consumo de energía ya que debido al uso inadecuado, anualmente se desperdicia una enorme cantidad de esta, también aseguran que el desperdicio de energía puede evitarse mediante una utilización eficiente de la misma y que se requieren soluciones inteligentes para certificar el uso adecuado de la energía pues una predicción del consumo es muy importante para conseguir un mantenimiento eficiente y reducir el impacto medioambiental; para ellos la gestión y eficiencia energética son las siguientes áreas cruciales donde se enfocarán las aplicaciones de Internet de las Cosas, IoT por sus siglas en inglés. Además realizan un resumen, en el que mencionan que se desarrollan muchas soluciones basadas en algoritmos de aprendizaje automático para la predicción del consumo de energía y que estos modelos utilizan datos históricos, que reflejan el comportamiento del proceso a modelar; las técnicas de aprendizaje automático que se han utilizado de forma redundante con fines de predicción son las redes neuronales artificiales, el sistema de inferencia neuro-difuso adaptativo (ANFIS), la máquina de vectores de soporte (SVM), la máquina de aprendizaje extremo (ELM), entre otras.

Saranj y Zolfaghari (2022) en su artículo titulado "The Electricity Consumption Forecast: Adopting a Hybrid Approach by Deep Learning and ARIMAX-G ARCH models" [8] proponen un enfoque híbrido para predecir el consumo de electricidad, de donde se obtiene como resultado que el empleo de transferencia de onda (WT por sus siglas en inglés) y bosque aleatorio (mejor conocido como Random Forest y sus siglas RF) para la descomposición de datos y la agregación no lineal de valores predichos, respectivamente, puede mejorar sustancialmente el poder de predicción de los modelos; ellos aseguran que pronosticar el consumo de electricidad ayuda a la industria a tomar mejores decisiones estratégicas.

Como mencionan Le, Thanh Vo, Vo, Hwang, Rho y Wook Baik en el artículo "Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM" [9], la predicción del consumo de energía eléctrica es una tarea esencial y compleja en un sistema inteligente de gestión de energía. En el mismo mencionan que la Predicción de Consumo de la Energía Eléctrica (EECP) desempeña un papel importante en la elaboración de una política nacional de desarrollo energético en Vietnam y es por eso que en este estudio se propuso el modelo llamado EECP-CBL, para la predicción del consumo de energía eléctrica, que utiliza la combinación de redes neuronales convolucionales (por sus siglas en inglés CNN) y redes neuronales de memoria de corto-largo plazo (por sus siglas en inglés Bi-LST. Los

investigadores además M). Se realizaron experimentos para indicar que EECF-CBL (el modelo propuesto) supera a otros modelos considerados de vanguardia para la predicción del consumo de energía eléctrica, en términos de diversas métricas de rendimiento como Error Cuadrático Medio (MSE), la Raíz del Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE) y el Error Porcentual Absoluto Medio (MAPE), así como el tiempo de procesamiento en diferentes configuraciones de marcos temporales.

Cárdenas, Ortega y Osma (2023) en su artículo titulado “Estrategias de predicción de consumo energético en edificaciones: una revisión” [10] hacen un repaso de las diferentes técnicas utilizadas en la predicción de consumo de energía eléctrica, en el estudio identifican cuatro enfoques de predicción: estadístico, ingeniería, basado en datos e híbrido; de estos, el modelo ARIMA es el más empleado dentro del enfoque estadístico, representando el 46,7 % de las investigaciones analizadas en este grupo. En segundo lugar, se encuentran los modelos SARIMA y SARIMAX, con un 20 % de utilización. Para la implementación de estas técnicas, Matlab es una de las herramientas más utilizadas, siendo empleado en el 26,7 % de los estudios revisados; por otro lado, mencionan que el enfoque estadístico en conjunto con el aprendizaje automático tuvo una representatividad del 58,3 % y para el enfoque de simulación simplificada y Machine Learning fue de 41,7 %. En un siguiente apartado los autores abordan que en cuanto al tipo de información a procesar, se recomienda utilizar el enfoque basado en datos o el estadístico si se tienen disponibles suficiente cantidad de datos históricos, sin embargo un enfoque híbrido podría utilizarse si se dispone de datos históricos y se desea compensar la incapacidad de los modelos estadísticos de tratar con los componentes no lineales de los datos, se puede usar el enfoque híbrido (estadístico+ Machine Learning). Para finalizar los autores identifican que la exactitud de las técnicas de predicción puede ser evaluada a través de indicadores de rendimiento y que a nivel general los más utilizados son RMSE y MAPE. Para Mejía y González y aq [11] es esencial anticipar la demanda eléctrica para equilibrar adecuadamente la oferta y la demanda, así como para gestionar eficientemente los aspectos de producción, distribución y consumo de electricidad en diversas escalas temporales, a corto, mediano y largo plazo, es por ello que en su trabajo titulado “Predicción del consumo de energía eléctrica residencial de la Región Cajamarca mediante modelos Holt – Winters”, aseguran que el modelo que presentó mejor ajuste para la serie de datos, es el modelo Holt-Winters aditivo, con el menor error de ajuste entre el modelo y la serie de datos.

Para Manobanda [12] la electricidad es una de las variables significativas para la industria, por lo que la anticipación de la demanda de la misma resulta crucial para garantizar el adecuado y eficiente desempeño de diversos activos, uno de los aspectos que el destaca en su investigación es que las redes neuronales artificiales son algoritmos de inteligencia artificial que mejoran la predicción de la demanda de energía eléctrica al emular el comportamiento no lineal asociado con la producción de fluido y esto posibilita a los planificadores reducir la incertidumbre en la toma de decisiones estratégicas.

2.2 MARCO TEÓRICO

2.2.1 Energía Eléctrica

La energía eléctrica es una forma de energía causada por el movimiento de electrones a través de un conductor, como un cable metálico. Se genera comúnmente mediante procesos como la conversión de energía mecánica a eléctrica en generadores o mediante reacciones químicas en pilas y baterías. Esta forma de energía es esencial en la vida moderna, utilizada para alimentar dispositivos electrónicos, iluminación, electrodomésticos, maquinaria industrial y más [13]. Para entender el concepto de energía eléctrica en términos comerciales, es útil considerar los términos indicados en la tabla 1:

Consumo	Cantidad total de energía eléctrica que una entidad (un hogar, una empresa, o una región) utiliza durante un periodo de tiempo determinado, generalmente medido en kilovatios-hora (kWh).
Demanda	Cantidad de energía eléctrica que se requiere en un momento específico o durante un periodo de tiempo determinado. Es una medida de la potencia necesaria para satisfacer las necesidades de los usuarios en una red eléctrica. La demanda no es una cantidad fija; varía a lo largo del tiempo, dependiendo de factores como la hora del día, el clima, las actividades económicas, y los patrones de consumo de los usuarios.
Energía Eléctrica	Es la capacidad de realizar trabajo mediante la electricidad. Se mide en julios (J) y es el producto de la potencia y el tiempo ($E = Pt$).
Eficiencia Energética	Cantidad de energía utilizada para realizar una tarea o producir un servicio en comparación con la cantidad de energía que realmente se necesita o se podría usar de manera óptima. En otras palabras, se refiere a

	la capacidad de obtener el máximo rendimiento posible con el menor consumo de energía.
Energía reactiva	Es aquella que no se convierte en trabajo útil, sino que es absorbida y devuelta por equipos como motores, transformadores, y otras cargas inductivas o capacitivas, influye en la eficiencia del sistema y el tamaño de la infraestructura requerida. Altos niveles de energía reactiva podrían estar asociados con cambios en la demanda de energía activa, ya que sistemas menos eficientes o con mayores pérdidas reactivas podrían requerir más energía activa para compensar.

Tabla 1.- Características claves de la Energía Eléctrica [13]

2.2.2 Gestión de la Energía Eléctrica

La gestión de energía eléctrica se refiere al conjunto de prácticas y estrategias implementadas para administrar eficientemente el consumo y suministro de energía eléctrica en un sistema, empresa, o instalación. Este campo abarca diversas áreas con el objetivo de optimizar el uso de la energía eléctrica, mejorar la eficiencia energética y reducir costos asociados [14].

2.2.3 Empresa Eléctrica Riobamba S.A.

La Empresa Eléctrica Riobamba S.A. (EERSA) opera dentro de la zona central del Ecuador, suministra electricidad a la provincia de Chimborazo, específicamente a los cantones de Riobamba, Chambo, Guano, Penipe, Guamote, Colta, Alausí, Chunchi, Pallatanga y Cumandá, lo hace tanto en áreas Rurales como urbanas, el trabajo de la EERSA involucra la transmisión, distribución y comercialización de energía eléctrica para satisfacer las necesidades de hogares, industrias y otros consumidores, concretamente los sectores definidos por la EERSA son residencial, comercial, industrial y agrícola. [15]

El presente proyecto investigativo centra su estudio en la zona urbana del cantón Riobamba, la electricidad en esta zona se suministra mediante cuatro transformadores principales, que a su vez se dividen en 6, 6, 5 y 4 alimentadores más los cuales no solo suministran a las diferentes parroquias urbanas del cantón así como tampoco son constantes, estos transformadores cuentan además con medidores que registran automáticamente el consumo, la demanda y la energía reactiva durante períodos de 15 minutos. [15]



Figura 1.- Mapa de la infraestructura eléctrica de la ciudad de Riobamba. [16]

En la figura 1, se muestra el mapa de la ciudad de Riobamba y sus alrededores, los diferentes colores muestran los alimentadores que dan servicio a cada zona, sin embargo las subestaciones no alimentan exclusivamente a la zona urbana de Riobamba, como se mencionó anteriormente, por lo que en general hacer una distribución exacta no es posible, puesto que para esto deberían existir datos de cada uno de los clientes con la misma frecuencia pero las mediciones de los clientes son mensuales.

2.2.4 Machine Learning

El machine learning, en español aprendizaje automático, es el estudio científico de modelos y algoritmos que sistemas computacionales utilizan para realizar tareas específicas sin la necesidad de ser programadas, es una rama de la inteligencia artificial. [17]

En el contexto de machine learning, estos algoritmos están presentes en varias aplicaciones de uso diario, como por ejemplo el procesamiento de imágenes, análisis predictivo y la minería de datos; y además pueden incluir una variedad de enfoques, como regresión, clasificación, agrupamiento (más conocido como clustering) y redes neuronales, entre otros. Estos algoritmos analizan datos de entrenamiento para identificar patrones y relaciones, y luego aplican ese conocimiento para hacer predicciones sobre datos nuevos o no vistos, en lugar de que el desarrollador programe manualmente cada escenario o excepción posible durante horas, simplemente se alimenta al algoritmo con una gran cantidad de datos. De esta manera, el algoritmo aprende y adquiere la capacidad de tomar decisiones en diversos casos sin necesidad de programación detallada. [17]

El machine learning, es usado para enseñar a las máquinas como tratar con datos de manera más eficiente; se basa en diferentes algoritmos para resolver problemas relacionados con datos. Los científicos de datos siempre señalan que no hay un único tipo de algoritmo que sirva para resolver todos los problemas o que sea mejor para resolver un problema en específico, el tipo de algoritmo a utilizarse depende del problema que se busca solucionar, el número de variables o el tipo de modelo que mejor se adapte [17]. Algunos algoritmos comúnmente usados en machine learning son los expuestos en la fFigura 2.- Algoritmos de Machine Learning.

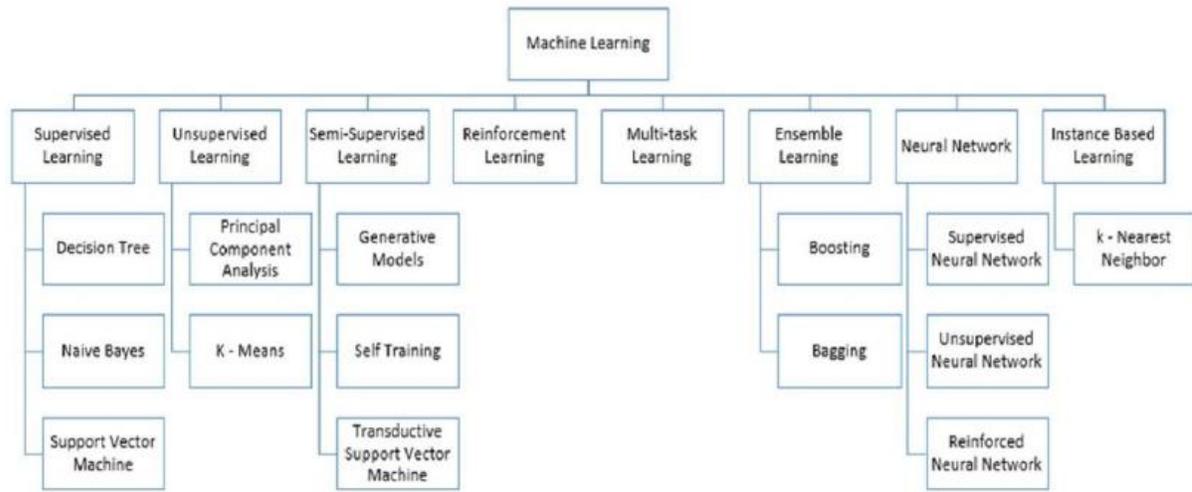


Figura 2.- Algoritmos de Machine Learning. [17]

2.2.5 Algoritmos de Machine Learning

Los algoritmos pueden diseñarse para realizar una amplia variedad de tareas, como ordenar una lista de números, realizar cálculos o buscar información. El principio de machine learning, consiste en alimentar algoritmos con la mayor cantidad posible de datos, esto con el objetivo de que el algoritmo “aprenda”, y no necesite de información por caso sino más bien que sea capaz de “saber” qué hacer en cualquier caso [18]. Con este contexto, se puede mencionar que existen dos tipos de aprendizaje.

- **Aprendizaje Supervisado**

Este tipo de aprendizaje consiste en alimentar al algoritmo con características y etiquetas, es decir darle las preguntas y respuestas posibles, de manera que en un futuro el algoritmo pueda hacer una predicción basándose en esta alimentación previa. Este tipo de aprendizaje se divide en algoritmos de clasificación y algoritmos de regresión, aunque estos términos más bien hacen referencia al tipo de entrenamiento de los algoritmos.

a) **Algoritmos de clasificación:** usando este algoritmo se espera como respuesta que los datos sean determinados como parte de un grupo, esto ya que el algoritmo debe encontrar patrones en los datos brindados y ubicarlos en un grupo; definiendo a la predicción como una variable, esta variable será un conjunto de estados discretos o categóricos [18]. Se puede considerar: (i) binaria; donde solo hay dos grupos, totalmente contrarios entre ellos, puede ser SI/NO, Azul/Rojo o Existe/No Existe, (ii) múltiple;, se puede escoger entre varios grupos, por ejemplo que producto se comprará en caso de tener varios como opciones, y finalmente (iii) ordenada; se da una escala y con base en las características se puede ubicar al dato en un punto de esta, por ejemplo al hablar de nivel de riesgo, se puede dar tres niveles como bajo, medio, alto [18].

b) **Algoritmos de regresión:** Se espera como respuesta un valor numérico, devuelve un valor específico, por ejemplo, el consumo de electricidad de una residencia en el siguiente mes, normalmente los resultados se muestran por medio de un gráfico de dispersión [18].

- **Aprendizaje no Supervisado**

En este caso el algoritmo solo recibe características y no etiquetas, es decir solamente es alimentado con preguntas, nunca con respuestas; por lo que el algoritmo determinará grupos asimilando características de los datos [18].

La elección del algoritmo muchas veces puede resultar difícil, porque existe una gran variedad de algoritmos con diferentes enfoques.

2.2.6 Modelos de Machine Learning

Es importante empezar definiendo un modelo, un modelo es la representación más simple de un sistema, los modelos ayudan a entender y predecir el comportamiento del sistema. Cuando se trata de machine learning, los algoritmos son habitualmente concebidos con el propósito de llevar a cabo tareas particulares, como la clasificación de datos o la realización de predicciones. En contraste, un modelo se define como una representación de un algoritmo de aprendizaje automático. Este modelo se nutre de un conjunto de datos durante su entrenamiento y luego se emplea para realizar predicciones o tomar decisiones basadas en dichos datos [18]. Los algoritmos de Machine Learning se podrían clasificar simplifícadamente para tres modelos:

Modelos Lineales	Modelos de árboles	Redes Neuronales Artificiales
Buscan encontrar una línea que se ajuste a la distribución	Son precisos, estables y más fáciles de interpretar, ya que	Buscan emular, en cierta medida, el funcionamiento

de puntos disponibles, incluyendo modelos populares como la regresión lineal (o mínimos cuadrados) y la logística (adaptada para problemas de clasificación).	construyen reglas de decisión representadas en forma de árbol. Esto permite abordar relaciones no lineales en la resolución de problemas.	del cerebro, donde millones de neuronas se interconectan para transmitirse mensajes. Esta reproducción de la actividad cerebral se destaca por sus habilidades cognitivas y capacidades de razonamiento.
---	---	--

Tabla 2.- Modelos de Machine Learning. [18]

- **Fases de Desarrollo**

Existen dos fases en el desarrollo de modelos para predicciones de eventos, estas son:

- Fase de Entrenamiento: Durante esta etapa, se dispone de una extensa cantidad de datos, de los cuales se selecciona una porción para instruir al algoritmo. Se le suministra esta información con el fin de que identifique los patrones esenciales y pueda realizar predicciones más adelante [18].
- Fase de Prueba: Los datos restantes se reservan para llevar a cabo pruebas. En esta etapa, se plantean preguntas al algoritmo para evaluar la precisión de sus respuestas. Este proceso permite determinar si el algoritmo está aprendiendo adecuadamente. En caso de discrepancias, se requiere la inclusión de más datos o la modificación del método utilizado. No obstante, si se observa que las respuestas son correctas en un rango del 80% al 90%, se puede concluir que el algoritmo ha alcanzado un nivel satisfactorio de aprendizaje y puede ser utilizado [18].

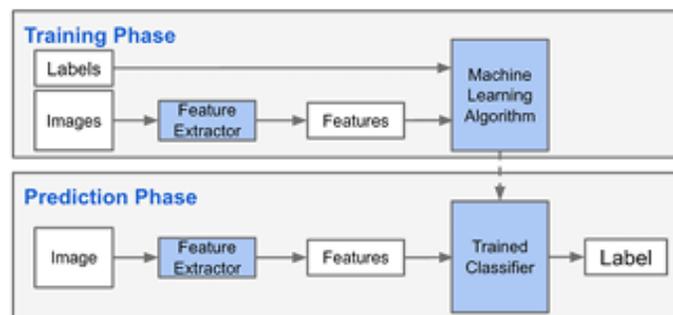


Figura 3.- Fases de modelado de algoritmos. [18]

2.2.7 Series Temporales

Las series temporales son conjuntos de datos que se recopilan o registran en intervalos regulares a lo largo del tiempo. En este tipo de datos, cada observación está asociada con un momento específico en el tiempo, y la disposición temporal de las observaciones es esencial

para comprender y analizar el conjunto de datos. Las series temporales se utilizan comúnmente en diversas disciplinas para estudiar la evolución de variables a lo largo del tiempo, es importante tomar en cuenta los siguientes conceptos para entender las series temporales; (i) orden temporal: las observaciones están organizadas en secuencia temporal, y el orden de las observaciones es fundamental para interpretar los datos; (ii) intervalos regulares: las observaciones se registran a intervalos fijos y regulares. Estos intervalos pueden ser un tiempo predefinido como horas, meses, años, etc.; (iii) tendencias y patrones: Las series temporales a menudo exhiben patrones, tendencias y estacionalidades. Estos patrones pueden ayudar a entender cómo una variable cambia en relación con el tiempo. Además las series temporales se utilizan en una variedad de campos, como la economía (para analizar el crecimiento económico), la meteorología (para prever el clima), la ingeniería (para monitorear el rendimiento de sistemas), la salud (para analizar datos médicos), entre otros [19].

- **Estacionalidad**

La estacionalidad es un comportamiento o patrón que se observa en una serie de tiempo, caracterizado por fluctuaciones regulares y repetitivas que se presentan en la serie de tiempo. Algunas características remarcables al hablar de estacionalidad son:

- (i) **Periodicidad:** los patrones repetidos ocurren una y otra vez en momentos determinados (como todos los días, todas las semanas, todos los meses).
- (ii) **Duración:** La duración de cada ciclo es fija y se repite. El uso de energía puede cambiar según el día de la semana o la hora del día.

Para analizar series temporales, se pueden aplicar diversas técnicas estadísticas y modelos, como redes neuronales artificiales, redes neuronales recurrentes, ARIMA, ARIMAX y otros, con el objetivo de comprender patrones pasados y hacer predicciones sobre futuras observaciones. [19]

2.2.8 Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA), o ANN por sus siglas en inglés, se pueden comparar con las estructuras neuronales biológicas en el cerebro humano. Estas redes se emplean principalmente en contextos donde resulta desafiante establecer restricciones lógicas directas, como en el reconocimiento de patrones y el análisis predictivo [20].

Las RNA son conceptualizadas como sistemas de procesamiento que se distinguen por la utilización de algoritmos que ejecutan operaciones simultáneas, lo que les habilita para realizar funciones como la identificación de patrones, la clasificación y la optimización [21].

Desde un punto de vista más centrado, una RNA se puede conceptualizar basándose en la definición de “grafo”, un grafo es un conjunto de nodos y las conexiones entre estos, que pueden clasificarse en grafos dirigidos y no dirigidos, donde los primeros hacen referencia a una conexión encaminada a un solo sentido, y los segundos a conexiones bidireccionales. En la Figura 4.- Estructura jerárquica de un sistema basado en RNA. [21] , se puede observar una representación tradicional de una RNA, esta guarda además una estrecha relación con la forma más habitual de la representación de un grafo.

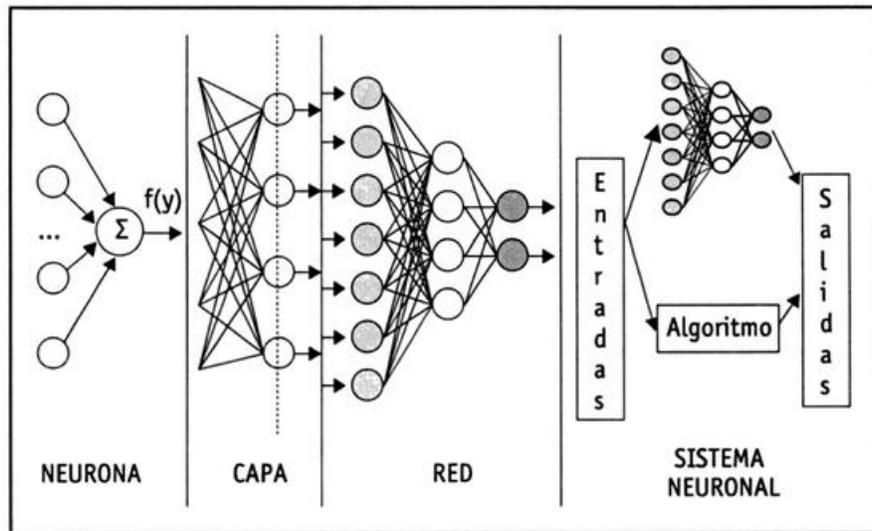


Figura 4.- Estructura jerárquica de un sistema basado en RNA. [21]

2.2.9 Redes Neuronales Recurrentes

Las redes neuronales recurrentes (RNN por sus siglas en inglés) son un tipo de arquitectura de redes neuronales que están diseñadas para manejar datos secuenciales o de series temporales. Las RNN tienen conexiones retroalimentadas que les permiten mantener y utilizar información de eventos anteriores en la secuencia. Esta capacidad las hace especialmente eficientes en tareas que involucran datos secuenciales, como el procesamiento del lenguaje natural, la traducción automática, la generación de texto, y también en problemas de predicción en series temporales, como el consumo de energía eléctrica [20].

En una RNN, cada nodo o unidad de la red tiene una memoria interna que retiene información de pasos de tiempo anteriores. Esto permite a las RNN aprender y recordar patrones y dependencias temporales en los datos. Sin embargo, las RNN pueden tener dificultades para retener información a largo plazo debido al problema de desvanecimiento del gradiente [20].

Para abordar este problema, han surgido variantes más avanzadas de RNN, como las redes neuronales de memoria de corto-largo plazo (mejor conocidas como LSTM por las siglas de Long Short-Term Memory), que incorporan mecanismos de puertas para controlar el flujo de

información como son la puerta de Olvido (Forget Gate), la cual decide qué información debe descartarse de la célula de memoria. Está definida por una función sigmoide que toma como entrada el estado anterior y la entrada actual, la puerta de entrada (Input Gate) que determina qué información nueva debe almacenarse en la célula de memoria. Una función sigmoide decide qué valores actualizar y una función tanh crea un vector de nuevos candidatos de memoria y la puerta de salida (Output Gate) la cual controla qué parte de la información en la célula de memoria se utilizará para la salida. Se aplica una función sigmoide para decidir qué parte del estado de la célula se usará; y otros mecanismos para mitigar el desvanecimiento del gradiente que se produce cuando una red tiene muchas capas, el valor del gradiente tiende a acercarse a cero porque estamos multiplicando repetidamente un valor pequeño. Este problema hace que las primeras capas de una red neuronal sean las más lentas y difíciles de entrenar, ya que el valor del gradiente utilizado para actualizarlas en cada iteración del entrenamiento es muy reducido; esto ocasiona otro inconveniente: si las primeras capas no se entrenan correctamente, el problema se transmite a las capas siguientes. Estas mejoras han hecho que las RNN sean más efectivas en la captura de patrones a largo plazo en datos secuenciales [20].

- **Redes LSTM**

Las redes LSTM (Long Short-Term Memory) son un tipo de arquitectura de RNN diseñada para superar las limitaciones de las RNN tradicionales en el manejo de dependencias a largo plazo en datos secuenciales. Fueron propuestas para abordar el problema del desvanecimiento del gradiente, que afecta la capacidad de las RNN para retener y utilizar información relevante de eventos pasados en secuencias temporales largas [6].

Las redes LSTM introducen una estructura más compleja en cada unidad de la red (célula LSTM), que incluye tres "puertas" fundamentales, puerta de olvido, puerta de entrada y puerta de salida [6].

Estas puertas permiten a las redes LSTM aprender y gestionar eficientemente la información a lo largo de secuencias temporales, facilitando la captura de patrones a largo plazo. Gracias a su capacidad para mantener y actualizar la información en la memoria de la célula, las redes LSTM son especialmente útiles en aplicaciones que involucran datos secuenciales complejos, como el procesamiento del lenguaje natural, la traducción automática, la generación de texto y la predicción en series temporales [6].

2.2.10 ARIMA y ARIMAX

ARIMA (AutoRegressive Integrated Moving Average) es un modelo estadístico que se utiliza para analizar y predecir datos temporales. Este modelo combina componentes autoregresivos (p), de integración (d) y de media móvil (q). Cada uno de estos componentes contiene aspectos específicos de la serie temporal, tales como: (i) componente autoregresivo (p): se refiere a la

dependencia lineal de la serie respecto a sus valores pasados. En otras palabras, la serie en un momento dado se relaciona con sus propios valores anteriores; (ii) componente de integración (d): indica el número de diferenciaciones necesarias para convertir la serie temporal en estacionaria, es decir, para hacer que la media y la varianza sean constantes en el tiempo; y (iii) componente de media móvil (q): se refiere a la dependencia lineal de la serie respecto a los términos de error pasados. Los términos de error anteriores se consideran para modelar las fluctuaciones no explicadas por las componentes autoregresivas [22].

La capacidad del modelo ARIMA para capturar tendencias y patrones en datos secuenciales lo convierte en una herramienta valiosa en la investigación científica y el análisis de datos. La ecuación 1 representa una forma simplificada del modelo ARIMA.

$$\begin{aligned}
 (Y_t = Y_t - Y_{t-d}) \quad Y_t \\
 = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots \\
 - \theta_q e_{t-q} + e_t
 \end{aligned} \tag{1}$$

Dónde:

- **Y_t** : Observación en el tiempo t de la serie temporal que estás analizando. Es el valor actual que deseas predecir o modelar.
- **c**: constante,
- **$\phi_1, \phi_2, \dots, \phi_p$** : Coeficientes de autorregresión de la parte autoregresiva del modelo, con "p" representando el orden de esta parte.
- **Y_{t-1}** : Valores anteriores de la serie temporal en los retrasos t-1, t-2, etc.
- **e_t** : Error o residuo en la predicción. Es la diferencia entre el valor observado real Y_t y el valor predicho por el modelo.
- **$\theta_1, \theta_2, \dots, \theta_q$** : Coeficientes de la parte de media móvil (Moving Average) del modelo, con "q" representando el orden de esta parte.

La ecuación (1) resume el proceso de modelado y predicción de series temporales teniendo en cuenta elementos como la dependencia del tiempo o el impacto de errores previos [22].

ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) es una extensión de ARIMA que incorpora variables exógenas en el modelo. Las variables exógenas son aquellas que no son parte de la serie temporal que se está analizando, pero que se cree que

tienen un impacto en ella. Al incluir estas variables en el modelo, ARIMAX permite capturar mejor la complejidad de las relaciones en los datos y mejorar la precisión de las predicciones al tener en cuenta factores externos [22]

Para hacer una comparación válida, mientras que ARIMA se centra exclusivamente en la serie temporal, ARIMAX amplía este enfoque al incorporar variables externas que pueden influir en el comportamiento de la serie temporal que se está modelando [22].

2.2.11 Modelos Híbridos

Dentro de esta categoría se incluyen técnicas que integran modelos estadísticos, como ARIMA, ARIMAX y SARIMAX, y algoritmos de aprendizaje automático, como las redes neuronales, con el objetivo de mejorar la precisión de las predicciones. [10]

El enfoque basado en datos tiene la capacidad de aprender y optimizar su desempeño a partir de la experiencia adquirida en la fase de entrenamiento. Este método es eficaz para resolver problemas no lineales y se sitúa en el ámbito de la computación y la inteligencia artificial, por otro lado, el enfoque estadístico emplea ecuaciones matemáticas para establecer relaciones lineales entre variables, permitiendo recopilar, organizar, analizar, interpretar y presentar datos con base en reglas predefinidas. Aunque ambos enfoques utilizan la regresión lineal fundamentada en la misma teoría, el enfoque basado en datos incorpora la regresión estadística y busca optimizar su precisión. [10]

2.2.12 Variables Exógenas

La teoría de las variables exógenas en modelos de predicción se basa en la idea de que factores externos tienen la capacidad de influir de manera notable en los resultados de un modelo predictivo. Estas variables, originadas fuera del proceso analizado, son independientes, pero se presume que pueden tener un impacto significativo en las predicciones del modelo. En síntesis, esta teoría sostiene que la inclusión de variables exógenas permite considerar factores adicionales que mejoran la precisión y la validez del modelo predictivo.

2.2.13 Métricas de Error

La validación de los modelos utilizados para el desarrollo del sistema, se la realiza con base en el cálculo de métricas de error, las métricas de RMSE, MAE y MAPE son las más populares en el contexto de series temporales.

- MAPE

El MAPE mide el porcentaje medio de error absoluto respecto al valor real. Por tanto, un MAPE del 10% indica que, en promedio, el modelo predice con un error de $\pm 10\%$ respecto al consumo real. En muchos estudios de series temporales energéticas, un MAPE menor a 10%

se considera muy bueno; de 10%–20% se considera bueno o aceptable y mayor a 20% comienza a ser poco confiable o deficiente. [23]

Empresas distribuidoras de energía y entornos operativos también consideran un MAPE < 10% como adecuado para toma de decisiones operativas y de planificación.

El MAPE tiene limitaciones cuando el consumo real se acerca a cero (divide por valores pequeños). Por eso se recomienda usarlo junto con MAE y RMSE para una visión completa del error. [23]

- MAE

El error absoluto medio es una métrica que mide el promedio de las diferencias absolutas entre las predicciones realizadas por un modelo y los valores observados, indicando en promedio cuanto se desvía el modelo respecto al valor real, sin considerar la dirección del error. [24]

El valor que puede tomar el MAE para que sea aceptable es bastante subjetivo debido a que este debe ser considerablemente menor a la media de los datos de, en este caso, consumo de energía.

- RMSE

La Raíz del Error Cuadrático Medio es una métrica utilizada para evaluar la precisión de modelos de predicción, como en el caso del consumo energético. Calcula la desviación promedio entre los valores reales y predichos, penalizando más los errores grandes. Se expresa en las mismas unidades de la variable predicha, lo que facilita su interpretación. Para validarlo, se calcula sobre un conjunto de datos de prueba y se compara con otros modelos. Un menor RMSE indica un modelo más preciso. [23]

En predicción de consumo eléctrico, el RMSE es necesario porque ayuda a identificar modelos confiables para la planificación y gestión energética. Su valor permite cuantificar qué tan bien se ajusta un modelo a la realidad. Si el RMSE representa menos del 10-15% del valor medio del consumo, se considera un resultado aceptable. Es útil tanto en pronósticos horarios como diarios. [23]

CAPÍTULO III. METODOLOGÍA.

3.1. TIPO DE INVESTIGACIÓN

Según el nivel de profundización en el objeto de estudio, la presente investigación es de tipo explicativa, debido a que para el análisis de los datos y su posterior intervención, se necesita comprender cuales son las variables que están interviniendo en el consumo de energía.

Según el tipo de datos empleados, a investigación es de tipo cuantitativo, porque los datos a estudiar son de carácter numérico y pertenecen al consumo de energía en la ciudad de Riobamba, además se realizarán análisis estadísticos de regresión temporal, para identificar patrones y relaciones entre el consumo de energía y variables que pueden afectar a este como la temperatura y de cambio estructural; para detectar momentos en los que los datos muestran un comportamiento diferente, como en el caso de festividades y eventos especiales.

Según el grado de manipulación de las variables, la investigación es de tipo cuasi-experimental, ya que en el desarrollo del sistema de predicción, los datos corresponderán únicamente a la ciudad de Riobamba, y se estudiarán con base en la realidad de la ciudad, tomando en cuenta variables únicas en la realidad planteada; además se estudiará diferentes algoritmos de manera que se pueda seleccionar los algoritmos y modelos más adecuados para la población de estudio.

Finalmente, según el objetivo, es de tipo aplicada, puesto que se centra en generar soluciones prácticas y directamente aplicables a situaciones o contextos específicos, como en este caso, donde se busca desarrollar un sistema de predicción que contribuya con la problemática que presenta la energía eléctrica y la gestión de la misma en la ciudad y país.

3.2. MÉTODOS DE INVESTIGACIÓN

El Método de Investigación en el que se basa el presente proyecto de Investigación es el método lógico de tipo sistémico, especialmente porque la predicción de consumo de energía eléctrica es un proceso que involucra múltiples factores y componentes interrelacionados. El enfoque sistémico es muy adecuado para este tipo de estudios porque permite comprender el consumo eléctrico no solo como una serie de datos temporales, sino como un fenómeno influido por variables externas, tendencias y relaciones entre diferentes elementos.

3.3 ENFOQUE DE LA INVESTIGACIÓN

La investigación tiene un enfoque predictivo, ya que el objetivo general es prever el comportamiento futuro del consumo de energía eléctrica en la ciudad de Riobamba, por lo que se utilizará modelos predictivos.

También tiene un enfoque analítico puesto que se realiza un análisis de los datos históricos, buscando identificar patrones, tendencias y correlaciones que ayuden a predecir el comportamiento futuro del consumo.

3.4. PROCESO METODOLÓGICO



Figura 5.- Proceso de la Metodología.

3.5. POBLACIÓN Y MUESTRA

3.5.1 Población

Para este estudio, se seleccionaron los datos de consumo de energía eléctrica correspondientes a los años 2021, 2022 y 2023. La elección de este periodo se justifica considerando que el año 2020 fue un año atípico por la pandemia global de COVID-19. Durante este periodo, las medidas de confinamiento y los cambios en las jornadas laborales (teletrabajo) tuvieron un impacto significativo pero no representativo para el estudio cuando se habla del consumo de energía eléctrica en general, el comportamiento del consumo de dicho año no refleja condiciones normales, cabe mencionar además que los hábitos de las personas cambiaron antes y después de este suceso sobre todo en el ámbito de consumo energético. Por ello, los hábitos de consumo anteriores a 2021 ya no son representativos, lo que podría sesgar los resultados del análisis. Por esta razón, se han excluido los datos de 2020 y años previos.

Por otro lado el año 2024 también fue excluido del estudio debido a que ha sido afectado por el estiaje que atraviesa la región y los consecuentes cortes de energía. Estas circunstancias han impactado en el consumo de energía, haciendo que los datos de dicho año no representen condiciones usuales para el sistema eléctrico.

La selección de los años 2021, 2022 y 2023 asegura que el análisis refleje patrones de consumo en condiciones más normales y comparables entre sí, minimizando la influencia de eventos externos que podrían sesgar los resultados.

La Empresa Eléctrica Riobamba S.A. cuenta con 14 subestaciones en la provincia de Chimborazo, 4 de ellas abastecen de energía eléctrica a la ciudad de Riobamba, aun cuando estas 4 subestaciones abastecen en su totalidad a la zona urbana, no son exclusivas, por lo que los datos obtenidos de consumo de energía eléctrica corresponden a la ciudad de Riobamba y sus alrededores. En estas subestaciones existen medidores que obtienen mediciones de demanda y consumo de energía eléctrica con una frecuencia de 15 minutos, por lo que es la mejor opción para el uso de algoritmos de predicción, puesto que estos necesitan la mayor cantidad de datos posibles para un mejor resultado.

De esta manera se han obtenido los datos de todos los días, durante las 24 horas, en intervalos de 15 minutos, de los años 2021, 2022 y 2023, lo que en total suman 105 120 datos con los que alimentar el algoritmo de predicción.

3.5.2 Muestra

La muestra analizada corresponde a la totalidad de la población por dos razones fundamentales. En primer lugar, los datos utilizados representan de manera válida los patrones de consumo eléctrico en la ciudad de Riobamba. En segundo lugar, al tratarse del desarrollo de algoritmos de predicción, es crucial contar con la mayor cantidad de datos posible para mejorar la precisión y robustez de los resultados obtenidos.

3.6 TÉCNICAS E INSTRUMENTOS

3.6.1 Técnicas de la Investigación

La principal técnica de investigación es el censo; mediante el cual se identificará patrones y tendencias al examinar las relaciones entre variables que podrían influir en el consumo de energía eléctrica en la ciudad de Riobamba.

Además se realiza el procesamiento de datos; con el objetivo de obtener datos de fuentes primarias que muestren la realidad de las condiciones de consumo de energía eléctrica en la ciudad de Riobamba para proceder a organizar estos datos y encontrar datos faltantes o datos atípicos y aplicar un método eficiente de procesamiento de los mismos.

Finalmente, se usa la experimentación para determinar el resultado del sistema predictivo que se pretende desarrollar.

3.6.2 Instrumentos de la Investigación

- Software R de análisis de datos para cálculos de medias, medianas y promedios después de recopilar datos sobre el consumo energético de la ciudad de Riobamba, se utilizarán apoyados de gráficas de línea, para analizar la información y determinar las variables que afectan al consumo y evaluar cómo afectan al mismo. Esto ayudará a inferir conclusiones válidas sobre el método adecuado que se adapte a la naturaleza de los datos y determinar conclusiones que ayuden a la eficiencia del sistema.
- Software MATLAB, permite modelar SARIMAX, RNN y LSTM, se utiliza el paquete de Econometrics Toolbox, que proporciona herramientas para modelar series temporales.
- Modelo Estadístico: SARIMAX, corresponde a un modelo basado en la metodología SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables). Este modelo permite incorporar variables exógenas, como temperatura y días festivos, con el objetivo de mejorar la precisión en la estimación del consumo energético.
- Modelo Híbrido: SARIMAX + Redes Neuronales, combinan el enfoque estadístico con técnicas de aprendizaje profundo. Este modelo busca mejorar la capacidad de generalización al identificar relaciones no lineales entre las variables de entrada y la demanda energética.
- Modelo Basado en Aprendizaje Profundo: Long Short-Term Memory - LSTM). Este modelo es capaz de capturar dependencias temporales más complejas, lo que permite mejorar la predicción del consumo energético a través del procesamiento secuencial de datos históricos.
- Registro del proceso de desarrollo, a través de una bitácora y comentarios en el código, donde se incluyen los cambios, problemas encontrados y soluciones implementadas, para evaluar la calidad del desarrollo y futuras mejoras.
- Registros de Observación: El principal instrumento de la investigación corresponde a los registros de observación, el sistema entregará resultados mediante gráficos estadísticos de línea, que serán observados y analizados por la investigadora.
- Métricas de Error, para evaluar la eficiencia del sistema.

3.7. OPERACIONALIZACIÓN DE VARIABLES

Variables	Tipo	Definición	Indicadores
Sistema predictivo	Dependiente	Utilizar algoritmos predictivos para a través de la utilización de datos históricos predecir el	Número de algoritmos que permitan desarrollar el sistema predictivo.

		consumo de energía eléctrica.	
Consumo de Energía Eléctrica	Independiente	Se cuantificará el consumo de energía eléctrica en kilovatios por hora (kWh) en diferentes sectores de la ciudad y a diferentes horas del día, a través del sistema predictivo.	Consumo de energía en kilovatios por hora (kWh) en diferentes sectores de la ciudad y a diferentes horas del día.

Tabla 3.- Operacionalización de Variables. Variables dependiente e independiente.

Es importante que en este apartado sea remarcado el hecho de que existe dependencia de variables exógenas en la investigación, ya que estos pueden afectar en gran medida la serie temporal que se estudia. Al pronosticar el uso de electricidad, elementos como el clima, las festividades o los días laborables pueden influir directamente en las tendencias de uso. Tomar en cuenta e incorporar estos factores permite capturar de manera más efectiva fluctuaciones y variaciones del consumo que no se observan en los registros históricos de la serie. Además, comprender la correlación entre estos factores externos permite generar ideas que mejoren la gestión de la energía y la formulación de estrategias para enfrentar la problemática presentada.

VARIABLES	TIPO	OPERACIONALIZACIÓN	INDICADORES
Día de la Semana	De Control	Registrar el día de la semana para considerar posibles variaciones en el consumo relacionadas con patrones laborales y actividades sociales.	Día de la semana al que pertenecen los datos.
Temperatura	Continua	Registrar los valores de temperatura que pueden influir en el consumo de energía eléctrica.	Valores de temperatura en °C diarios en la ciudad de Riobamba durante el período de estudio.
Festividades y Eventos Especiales	Moderadora	Identificar y registrar la presencia de	Registro de fechas donde el consumo de

		festividades o eventos especiales que puedan influir en el consumo de energía en la ciudad, y puede causar valores atípicos.	energía tuvo un consumo no habitual debido a festividades o feriados.
--	--	--	---

Tabla 4.- Operacionalización de Variables. Variables exógenas.

3.8 PRE PROCESAMIENTO DE DATOS

3.8.1 Limpieza de Datos

- **Detección de Valores Faltantes**

Se realiza una breve revisión manual de los datos entregados por la EERSA, se identifica la ausencia de un valor correspondiente al 31/11/2021 a las 23:45, considerando que es el único valor faltante, se reemplaza obteniendo la media de los datos anteriores.

- **Tratamiento de valores atípicos (outliers)**

a. Clustering

El análisis de Clustering es una herramienta poderosa para identificar outliers en el consumo de energía cuando se consideran variables exógenas como la temperatura, el día de la semana y los días festivos. A diferencia de métodos más simples como el análisis basado en la desviación estándar, el Clustering permite agrupar los datos según patrones complejos que surgen de la interacción de múltiples variables. El consumo de energía puede variar significativamente no solo por la temperatura, sino también dependiendo de si es un día laboral o un fin de semana. El Clustering detecta estos patrones subyacentes, creando grupos de datos similares. Los puntos que no se ajustan bien a ningún clúster, es decir, aquellos que están alejados de los centroides o quedan aislados, pueden ser identificados como outliers. El análisis de clústeres es muy útil porque no asume una distribución específica de los datos y puede adaptarse a las variaciones estacionales o contextuales, ofreciendo una detección más precisa y robusta de anomalías en el consumo energético.

Para el desarrollo del mismo se comienza por encontrar la variable k, que indica el número de clústeres recomendados para el análisis, para lo mismo se hace uso del software de estadística R, y del método del codo, obteniendo la Figura 6:

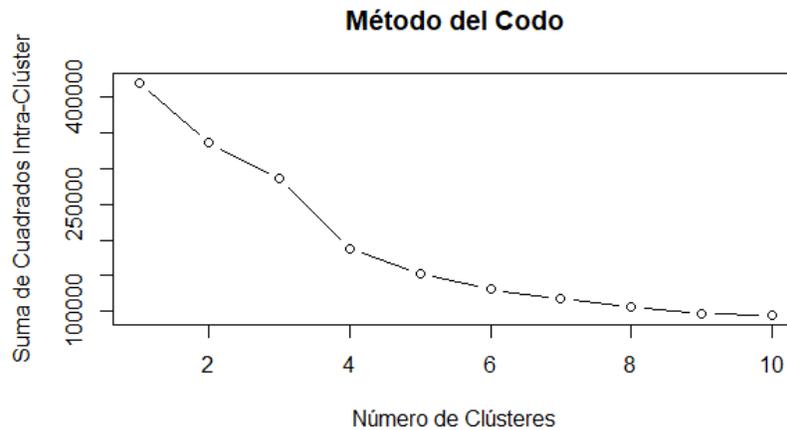


Figura 6.- Método del Codo para identificar el número adecuado de clústeres.

Con base en la teoría previamente explicada sobre el análisis del codo, podemos identificar que el número necesario de clústeres está entre 4 y 6. Una vez obtenido este valor se realiza el análisis k-means para identificar outliers. En la Figura 7 se observa el resultado:

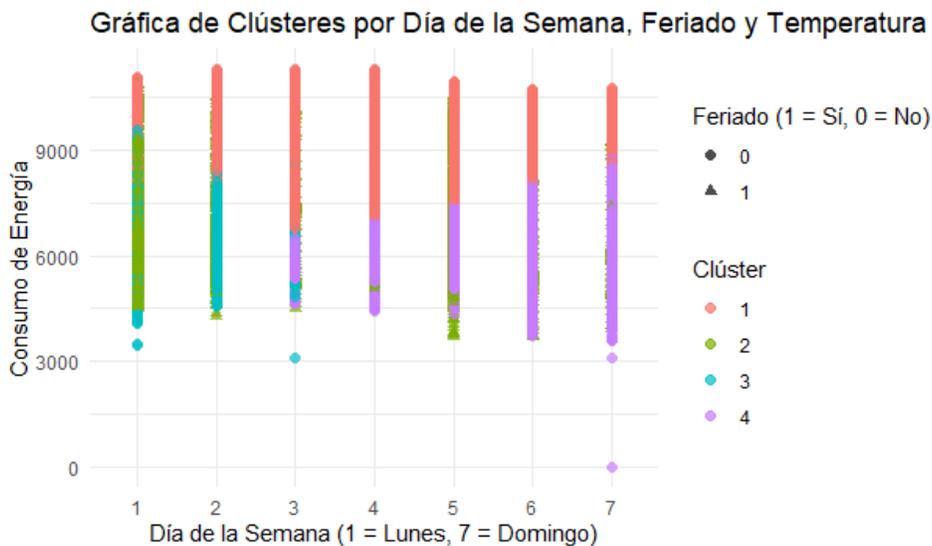


Figura 7.-Gráfica de Clústeres tomando en cuenta las variables exógenas.

La gráfica presentada permite evidenciar que existen puntos fuera de los rangos que se considera que pertenecen a un grupo o clúster específico, los cuales fueron determinados tomando en cuenta las variables exógenas que han sido señaladas durante la investigación.

Se acompaña la Figura 7 con la Figura 8, en la cual se observa un diagrama de caja o boxplot, mientras el análisis de clúster agrupa datos similares, la gráfica de caja permite observar la

variabilidad interna de cada grupo y resaltar posibles anomalías (outliers) dentro de esos grupos.

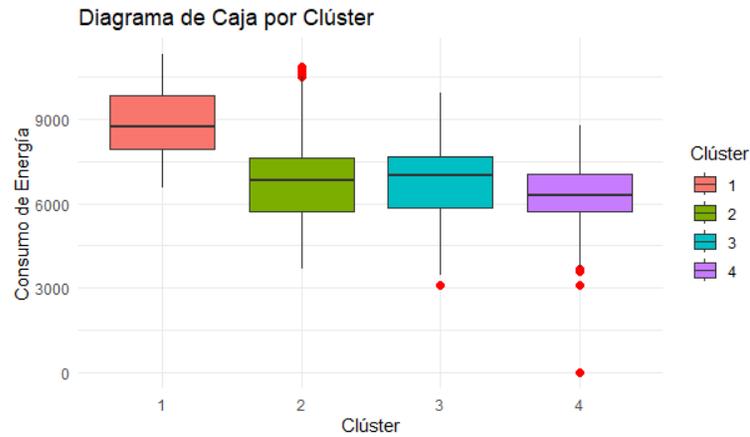


Figura 8.- Diagrama de caja o boxplot por clústeres.

Gracias a estas herramientas estadísticas se obtiene la tabla de la Figura 9, en la cual se identifica cada uno de los outliers que el modelo ha encontrado:

Consumo	Temperatura	Condicion
<dbl>	<dbl>	<dbl>
10524.	12.7	1
10535.	13.6	1
10543.	13.6	1
10535.	12.6	1
10593.	12.6	1
10562.	12.6	1
10643.	13.6	1
10745.	13.6	1
10728.	13.6	1
10637.	13.6	1
10564.	13.6	1
10506.	14.1	1
10587.	12.4	1
10658.	12.4	1
10654.	12.4	1
10616.	12.4	1
10562.	11.5	1
10555.	11.5	1
10515.	11.5	1
10748.	14.6	1
10853.	14.6	1
10843.	14.6	1
10817.	14.6	1
10717.	14.6	1
10591.	14.6	1
3112.	11.3	0
3601.	12.7	0
3596.	12.7	0
3641.	12.5	0
3664.	12.5	0
3082.	12.5	0
0	12.5	0

Figura 9.- Tabla de outliers generada por el método de Clustering.

Pese a existir variedad de modelos automáticos para determinar outliers, como los presentados, complementarlos con una revisión manual es necesario para proporcionar precisión contextual y garantizar que los datos no solo sean estadísticamente válidos, sino también operativamente consistentes pues si los outliers son datos reales y se eliminan o reemplazan sin justificación, puede haber distorsiones en los análisis posteriores, especialmente en modelos predictivos, donde datos válidos mal clasificados pueden afectar la precisión del modelo, además reemplazar un dato real con un valor artificial puede reducir la calidad de la base de datos y generar conclusiones erróneas.

A continuación en la Tabla 5 se presentan los outliers en los cuales la investigadora ha intervenido, los demás outliers después de una revisión manual complementaria se han identificado como operativamente consistentes:

Fecha	Consumo	Consumo corregido	Observación
28/6/2023 18:30:00	3112	9211,52	El dato es reemplazado por la media de los datos 15 minutos antes y 15 minutos después.
26/3/2023 9:30:00	3082	6367,00	El dato es reemplazado por la media del clúster al que pertenece.
26/3/2023 9:45:00	0,0	6367,00	El dato es reemplazado por la media del clúster al que pertenece.

Tabla 5.- Outliers modificados en la base de datos.

- **Corrección de errores**

Después del análisis de outliers, la corrección de errores resulta más sencilla sin embargo se revisa nuevamente de manera manual los datos uno a uno, para identificar errores tipográficos, inconsistencias evidentes y datos mal registrados; en este caso no se encontraron.

3.8.2 Reducción de Dimensionalidad

- Selección de características

a. Consumo

En la Tabla 6 se observa la forma en que los datos fueron entregados a la investigadora para su procesamiento.

Time stamp	SUB_1.TOTAL1 Reactive Energy Into the Load Interval (kVArh)	SUB_1.TOTAL1 Real Energy Into the Load Interval (kWh)	SUB_2.TOTAL2 Reactive Energy Into the Load Interval (kVArh)	SUB_2.TOTAL2 Real Energy Into the Load Interval (kWh)	SUB_3.TOTAL3 Reactive Energy Into the Load Interval (kVArh)	SUB_3.TOTAL3 Real Energy Into the Load Interval (kWh)	SUB_4.TOTAL4 Reactive Energy Into the Load Interval (kVArh)	SUB_4.TOTAL4 Real Energy Into the Load Interval (kWh)	SUB_4.TOTAL4 Reactive Energy Out of the Load Interval (kVArh)	SUB_4.TOTAL4 Real Energy Out of the Load Interval (kWh)	SUB_ALAO.TOTAL_ALAO Reactive Energy Into the Load Interval (kVArh)	SUB_ALAO.TOTAL_ALAO Reactive Energy Out of the Load Interval (kVArh)	SUB_ALAO.TOTAL_ALAO Real Energy Into the Load Interval (kWh)	SUB_ALAO.TOTAL_ALAO Real Energy Out of the Load Interval (kWh)	SUB_ALAUSI.TOTAL_ALAUSI Reactive Energy Into the Load Interval (kVArh)	SUB_ALAUSI.TOTAL_ALAUSI Reactive Energy Out of the Load Interval (kVArh)	SUB_ALAUSI.TOTAL_ALAUSI Real Energy Into the Load Interval (kWh)	SUB_ALAUSI.TOTAL_ALAUSI Real Energy Out of the Load Interval (kWh)
1/4/2021 00:00	410,01	1,497,73	359,87	1,996,13	397,73	1,418,67	341,76	1,432,51	0,00	0,00	33,69	0,00	346,05	0,00	57,17	0,00	236,63	0,00
1/4/2021 01:15	420,12	1,451,91	352,18	1,941,72	397,23	1,369,49	348,32	1,387,66	0,00	0,00	34,13	0,00	338,72	0,00	60,78	0,00	231,60	0,00
1/4/2021 02:30	424,87	1,417,42	364,60	1,931,72	396,92	1,328,97	348,10	1,353,13	0,00	0,00	35,32	0,00	330,78	0,00	61,45	0,00	230,61	0,00
1/4/2021 03:45	417,78	1,374,79	350,25	1,862,51	390,07	1,299,32	341,18	1,317,54	0,00	0,00	35,12	0,00	328,72	0,00	58,85	0,00	223,97	0,00
1/4/2021 04:00	413,76	1,344,88	349,03	1,824,56	387,92	1,273,84	337,59	1,291,75	0,00	0,00	34,71	0,00	325,53	0,00	56,56	0,00	218,17	0,00

21																	
1:																	
00:																	
:0																	
0																	
1/																	
4/																	
20																	
21	415	1.3	337	1.8	378	1.2	334	1.2	0,0	33,88	0,00	323,50	0,00	54,84	0,00	213,88	0,00
1:	,36	24,	,39	00,	,51	47,	,83	72,									
15		70		27		53		69									
:0																	
0																	

...

Tabla 6.- Formato de la base de datos proporcionada por la fuente.

Como se observa existen varios campos que no son del interés de la investigación por lo que han sido eliminados; además que no existe un compilado de los 4 transformadores de interés, por lo que fue necesario realizar uno donde se simplifica el consumo a uno y no cuatro como se observa en la tabla; finalmente, se escogen las características con las que se va a trabajar que corresponden a ‘SUB_1.TOTAL1 Real Energy Into the Load Interval (kWh)’, ‘SUB_2.TOTAL2 Real Energy Into the Load Interval (kWh)’, ‘SUB_3.TOTAL3 Real Energy Into the Load Interval (kWh)’ y ‘SUB_4.TOTAL4 Real Energy Into the Load Interval (kWh)’.

3.8.3 Integración y Fusión de Datos

- **Integración de Variables**

a. Temperatura

Se incorporó la variable de temperatura como un factor exógeno para mejorar la precisión del modelo de predicción de consumo energético. Los datos de temperatura fueron obtenidos de fuentes en línea confiables, lo que permitió asociar las variaciones climáticas con el comportamiento del consumo eléctrico en la ciudad de Riobamba. La integración de esta variable busca capturar el posible impacto de las fluctuaciones de temperatura sobre el consumo energético.

b. Día de la semana y condición de feriado

Para enriquecer la investigación, se incluyeron las variables día de la semana y condición de feriado como factores adicionales que podrían influir en el consumo energético. Estas variables fueron incorporadas para capturar patrones y variaciones asociados al comportamiento diario y los efectos de los días no laborables en el consumo. La información sobre los días feriados se recopiló de fuentes oficiales y calendarios locales, y se codificó para reflejar su impacto en el análisis. Esto permite un mejor análisis del consumo energético, al considerar cómo las rutinas semanales y los días festivos modifican las tendencias de uso.

- **Fusión de Datos**

Finalmente se consolida una base de datos, donde se observa la fecha y hora del consumo, el total de consumo en kWh, la temperatura promedio del día, el día de la semana al que corresponde el consumo y la condición de feriado. Se puede acceder al repositorio <https://github.com/Vpvicpar/Codigos-Predictivos-EERSA> para una revisión de la base de datos consolidada, en el archivo “datoscsv”.

3.8.4 División del Conjunto de Datos

Es necesaria una división en subconjuntos de entrenamiento, validación y prueba para crear un modelo que sea capaz de generalizar y evaluar correctamente su desempeño; tomando en cuenta que se tiene 105120 datos correspondientes a tres años de consumo de energía eléctrica, se propone la siguiente división temporal de datos:

- **Propuesta de División de Datos**

La propuesta a continuación se presenta a razón de que el orden de los datos importa, por lo que mantener la secuencia temporal es necesario pues el modelo no debe ver el futuro, es decir no debe conocer los datos del futuro durante el entrenamiento y la validación así se consigue que el modelo sea estable, pues al tener gran cantidad de datos como en este caso, dividirlos de esta manera proporciona suficientes datos para cada subconjunto.

- 1) Entrenamiento 60% → 63072 datos
- 2) Validación 20% → 26276 datos
- 3) Prueba 20% → 26276 datos

3.9. ESTACIONALIDAD DE LOS DATOS

En el presente proyecto se trabaja con el total de la población como muestra, la población es de 105120 que corresponde a 96 datos diarios en intervalos de 15 minutos en los tres años de análisis, dado que se trabaja con el total de la población, se opta por la compresión de los datos, es decir, se realizará el cálculo del promedio del consumo en cada intervalo horario con el objetivo de obtener una única gráfica, esto no implica pérdida de representatividad, sino solo un cambio en la granularidad de los datos.

En la figura 10 y 11, se observa el comportamiento del consumo, por lo que se determina que existe estacionalidad diaria.

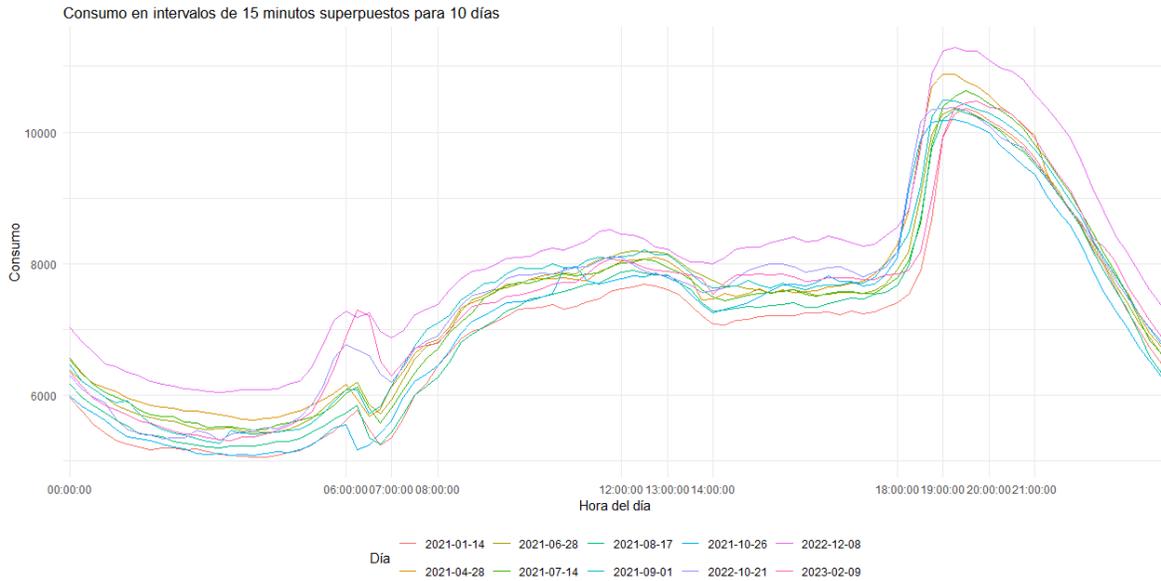


Figura 10.- Consumo de 10 días aleatorios en el conjunto de datos.

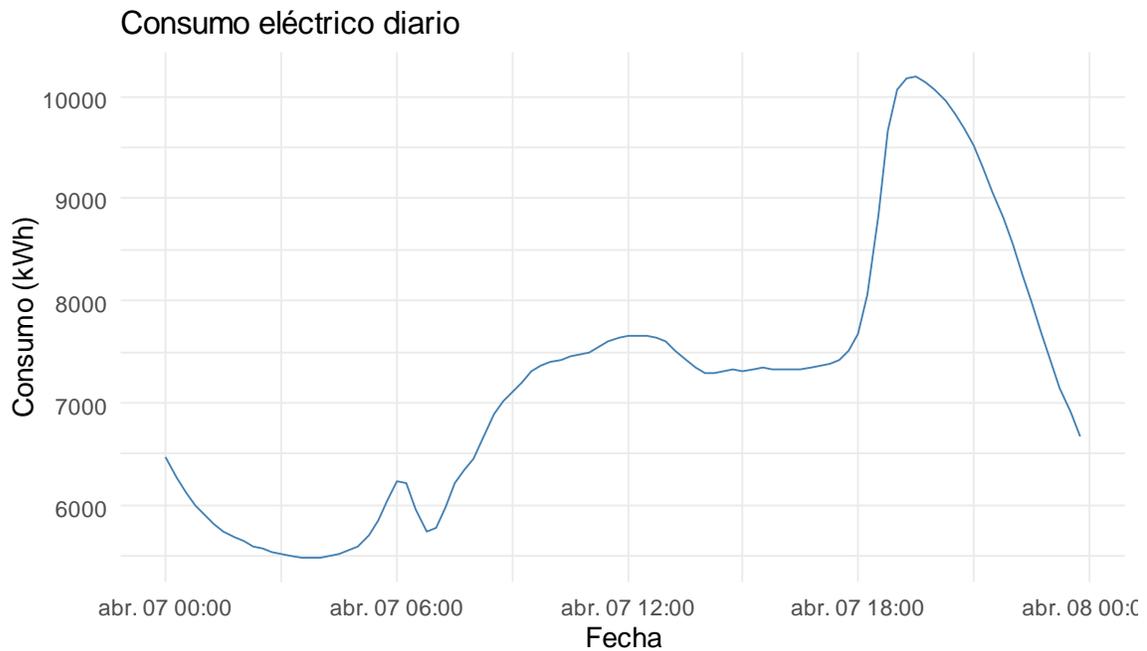


Figura 11.- Consumo promedio diario de energía eléctrica en la ciudad de Riobamba en kWh.

En general se observa que diariamente el consumo de energía eléctrica sigue un patrón, donde el consumo tiene un pico a las 06:00 que desciende alrededor de las 07:00 para ascender nuevamente a las 08:00 hasta medio día donde nuevamente desciende hasta alrededor de las 14:00 donde asciende sutilmente hasta las 18:00 donde se evidencia un crecimiento significativo alrededor de las 19:00 hasta las 21:00 donde nuevamente el pico decrece

significativamente, cabe aclarar que el consumo nunca se acerca a 0. Con base en estos resultados y evidenciando que pueden existir otros factores que influyan en el consumo de energía eléctrica se realizan más pruebas.

3.10 RELACIÓN ENTRE EL CONSUMO Y LAS VARIABLES EXÓGENAS (DÍAS DE LA SEMANA, FERIADOS Y TEMPERATURA)

3.10.1 Relación entre el Consumo y la Temperatura

El gráfico de dispersión de la Figura 12 contribuye a determinar si existe una tendencia clara en los datos, que puede ser, lineal, cuadrática o de otro tipo. Lo importante es mencionar que una prueba de regresión lineal o polinómica solo será válida si la tendencia es clara.

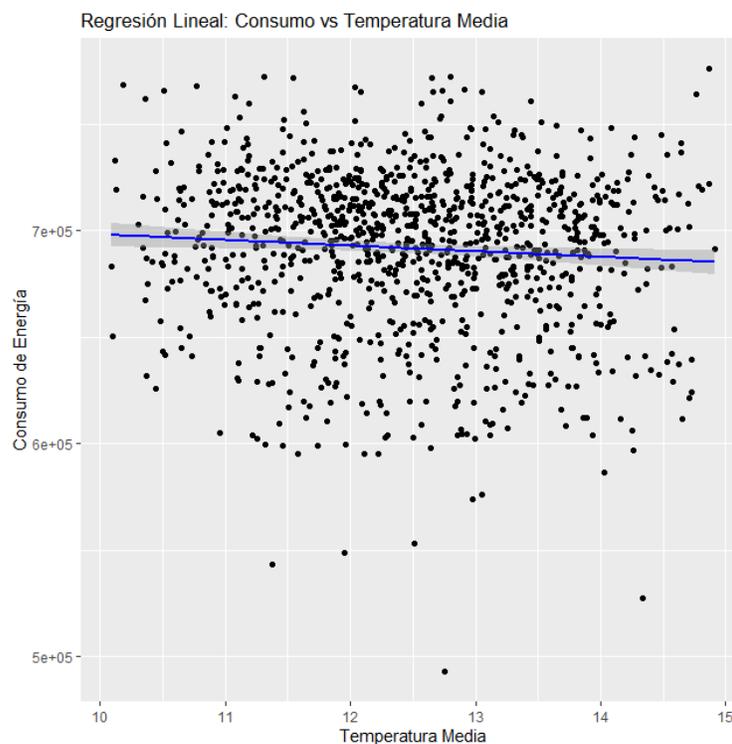


Figura 12.- Gráfica de dispersión de los datos de Consumo vs Temperatura.

La gráfica de dispersión no es clara, no se puede determinar que exista una tendencia por lo que no es posible hacer pruebas de regresión para determinar la relación entre estas variables, por lo que se utiliza el coeficiente de correlación de Pearson r .

Se obtiene un valor de -0.0734057 , se sabe que $r \in [-1,1]$, y que un valor negativo significa una relación negativa así como que un coeficiente muy cercano a 0 explica que existe poca relación lineal, tomando en cuenta que hablamos de un censo, este valor describe por completo la fuerza de la relación.

El valor en general indica que, en promedio, cuando la temperatura media aumenta el consumo de energía eléctrica tiende a disminuir, ligeramente.

3.10.2 Relación entre el Consumo y los días de la semana

TipoDia	MediaConsumo
<chr>	<dbl>
1 FinDeSemana	<u>654834.</u>
2 Semana	<u>706460.</u>

Figura 13.- Media de Consumo de los días entre semana vs los fines de semana.

Los resultados determinan que la diferencia entre las medias de consumo de los días entre semana y fines de semana es de 51626.6 KWh, es decir que los fines de semana se utiliza en promedio 51626.6 KWh menos que durante los días entre semana, por lo que la condición de día entre semana o fin de semana es, con base en el censo, una variable exógena dentro del sistema.

3.10.3 Relación entre el Consumo y la condición de feriado

feriado	MediaConsumo
<chr>	<dbl>
1 no	<u>692741.</u>
2 si	<u>664467.</u>

Figura 14.- Media de Consumo vs condición de feriado.

En la Figura 14 se observa los valores de la media de consumo de los días feriados y los días que no lo son, la diferencia de consumo entre ellas es de 28273.44 kWh, los días no feriados independientemente de que sean o no entre semana se consume en promedio 28273.44 kWh más que los días feriados. Lo que determina que la condición de feriado es una variable exógena dentro del sistema.

3.11. ENTRENAMIENTO DEL MODELO

Para desarrollar el sistema de predicción se ha escogido el software de MATLAB, debido a que el mencionado programa ofrece funciones y herramientas específicas para el manejo y análisis de series temporales, como el Time Series Toolbox, mismo que es ideal para trabajar con datos de consumo eléctrico, ya que permite manejar y procesar datos ordenados cronológicamente de manera eficiente, realizar interpolaciones, descomposiciones y visualizar datos temporalmente, además facilita el ajuste de modelos estadísticos a datos de series temporales, incluido el modelo ARIMA, Redes Neuronales y LSTM; además permite integrar código de otros lenguajes como Python y tiene funciones de exportación de datos compatibles con Excel lo que permite combinar diferentes herramientas.

Para identificar el modelo que mejor se adapte a la naturaleza de los datos de consumo eléctrico, se realizaron pruebas de entrenamiento con diferentes enfoques de series temporales. Se implementaron modelos autorregresivos como ARIMA y ARIMAX, que consideran únicamente la estructura temporal y la posible influencia de variables exógenas. Adicionalmente, se evaluaron modelos estacionales como SARIMAX, que incorpora patrones cíclicos propios del consumo energético. Posteriormente, se exploraron enfoques híbridos mediante la combinación de SARIMAX con Redes Neuronales, así como arquitecturas de aprendizaje profundo basadas en redes LSTM (Long Short-Term Memory). La selección del modelo óptimo se basó en métricas de desempeño como el error absoluto medio (MAE), el error cuadrático medio (MSE) y el error porcentual absoluto medio (MAPE), permitiendo determinar la mejor alternativa para la predicción del consumo eléctrico en Riobamba.

En esta sección se presentan los modelos de predicción evaluados en este estudio. Se inicia con los modelos estadísticos tradicionales como ARIMA y ARIMAX, los cuales permiten capturar patrones temporales y considerar variables externas en la predicción. Luego, se describen los modelos estacionales como SARIMAX, diseñados para datos con variaciones cíclicas. Posteriormente, se exploran enfoques más avanzados basados en aprendizaje automático, incluyendo la combinación de SARIMAX con Redes Neuronales y el uso de redes neuronales recurrentes del tipo LSTM, las cuales destacan por su capacidad para modelar secuencias de datos con dependencias a largo plazo.

Durante la implementación de los modelos, se llevó un registro de desarrollo, que se puede encontrar en la sección de Anexos (Anexo 1), además se emplearon comentarios en el código con el fin de facilitar la comprensión y mantenimiento del mismo. Estos comentarios explican el propósito de funciones clave y la configuración de los parámetros utilizados en los modelos predictivos. La documentación de los códigos se encuentra disponible en el repositorio del proyecto. Para acceder a los archivos y revisar la documentación completa se puede acceder al siguiente enlace: <https://github.com/Vpvicpar/Codigos-Predictivos-EERSA>.

3.11.1 Desarrollo del modelo SARIMAX

Para modelar ARIMA, ARIMAX y SARIMAX se necesitan los parámetros p , q , d , P , Q , D , s ; los cuales se obtienen de la serie temporal. Los modelos ARIMA, ARIMAX y SARIMAX son implementados utilizando las funcionalidades disponibles para modelos ARIMA, extendidos con términos estacionales y variables exógenas.

- **Parámetros p , d , q**

En el modelo ARIMA se busca los parámetros p , d , q que corresponden a las órdenes del modelo.

Para encontrar el parámetro p u orden de la parte autorregresiva, se realiza la observación del gráfico de la función de auto correlación parcial (PACF), el valor de p corresponde al último rezago significativo antes de que la PACF se reduzca a valores cercanos a 0.

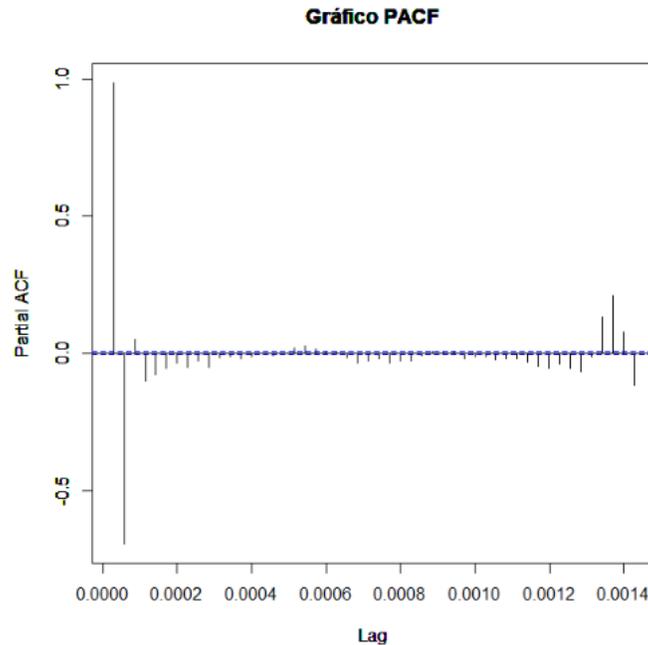


Figura 15.- Gráfica PACF para determinar parámetros ARIMA.

En la Figura 15, se observa que después del primer rezago, las correlaciones parciales parecen caer cerca de 0 y están dentro del rango de significancia. Esto sugiere que no hay dependencia significativa entre los valores posteriores de la serie y el valor actual después del primer rezago, por lo que el orden autorregresivo $p = 0$ podría ser el adecuado.

Después, para obtener el parámetro q , el cual indica el número de términos de error pasados (rezagos del error) que se utilizan en el modelo para predecir el valor actual; se puede observar el gráfico de la Función de Autocorrelación (ACF). Se estima que el valor de q es el último rezago significativo antes de que la ACF disminuya a valores cercanos a cero.

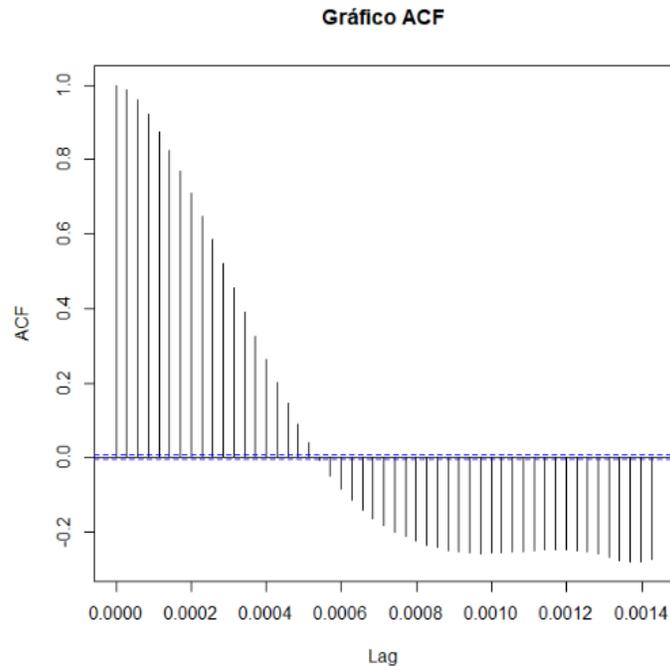


Figura 16.- Gráfica ACF para determinar parámetros ARIMA.

En el gráfico ACF de la Figura 16, la correlación muestra una disminución gradual en lugar de un corte brusco, lo que significa que no se evidencia un patrón claro de corte en la ACF, por lo que un término de media móvil complejo no es necesario y $q = 0$ podría ser adecuado.

Finalmente, el parámetro d de diferenciación, representa el número de veces que se diferencia la serie para hacerla estacionaria. Se puede obtener basándose en la prueba de estacionalidad realizada anteriormente, la cual resolvió que la serie temporal presentada es estacional, por tanto tiene tendencia, y es necesario una componente de diferenciación, se puede utilizar $d=1$. Es importante aun cuando la serie temporal es estacional ya que estos parámetros describen la parte no estacional de la serie temporal.

- **Parámetros P, D, Q, s**

El modelo SARIMA añade términos de autorregresión, diferenciación e integración que capturan la estacionalidad de la serie temporal. Estos términos son representados por parámetros estacionales (P, D, Q, s).

El parámetro s , o estacionalidad, representa el número de puntos por período estacional, una vez determinada la estacionalidad diaria de la serie temporal, se determina que $s=96$, ya que representa un ciclo diario, y también se define a la estacionalidad diaria como la más importante dentro de la serie temporal.

Para determinar los parámetros D, Q y P se reutilizan la Figura 15 y Figura 16; la diferenciación estacional D será de 0, esta se usa cuando hay un patrón estacional que debe eliminarse para estabilizar la serie. Para determinar el orden autorregresivo estacional P se revisa la gráfica PACF, esta muestra un pico significativo en el primer rezago estacional y luego cae a valores no significativos, lo que indica que solo el primer término estacional autorregresivo es necesario para capturar la dinámica de la serie, por lo que P tendrá un valor de 0, esto indica que los valores actuales están relacionados principalmente con los valores del día anterior; finalmente el valor de Q se determina observando si el gráfico ACF tiene un pico significativo en el primer retraso estacional, ya que lo tiene y que la serie temporal se determinó estacional, se puede considerar $Q=0$.

En la Tabla 7 se presenta un resumen de los parámetros establecidos teóricamente y los que se utilizaron en la predicción final:

Parámetros	p	q	d	s	P	D	Q
Teórico	0	0	1	96	0	0	0
Práctico	0	1	0	96	0	0	0

Tabla 7.- Parámetros para los modelos SARIMAX y SARIMAX+RNN.

- **Variables Exógenas**

Los modelos ARIMAX y SARIMAX usan además de los parámetros ya mencionados, variables exógenas, estas se incorporan para mejorar la capacidad predictiva del modelo de consumo eléctrico. Estas variables han sido obtenidas y procesadas de la siguiente manera:

- **Día de la semana (1-7):** Esta variable ha sido generada mediante fórmulas en Excel a partir de la fecha correspondiente a cada registro. Se ha asignado un valor numérico del 1 al 7, donde 1 representa el lunes y 7 el domingo, lo que permite identificar patrones de consumo asociados a la periodicidad semanal.
- **Condición de feriado (1 o 0):** Debido a la ausencia de una fuente automatizada para esta información, se ha ingresado manualmente en la base de datos. Se ha asignado un valor de 1 a los días festivos y 0 a los días laborables, con el fin de evaluar el impacto de los feriados en la demanda de energía.
- **Temperatura media:** Esta variable ha sido extraída de una base de datos en línea que proporciona registros históricos de temperatura. Se ha asociado la temperatura promedio diaria a cada observación correspondiente en la serie de consumo eléctrico,

permitiendo analizar la posible relación entre la demanda de energía y las condiciones climáticas.

Estas variables exógenas complementan la variable dependiente de consumo, facilitando la identificación de patrones y mejorando la precisión del modelo predictivo. En todos los casos, las variables exógenas han sido exportadas a MATLAB para su procesamiento e integración en el modelo de predicción. Cada una de ellas ha sido identificada con el término X, permitiendo su correcta manipulación y análisis dentro del entorno de programación. Esto garantiza la coherencia de los datos y facilita su uso en las distintas etapas del desarrollo del sistema de predicción de consumo eléctrico.

- **Predicción y Evaluación del Modelo**

En la Figura 17 se observa la predicción realizada por el modelo SARIMAX, que se encuentra disponible en el repositorio del proyecto, con los parámetros (0,1,0) (0,0,0) y una estacionalidad de 96, que representa un patrón diario con 96 observaciones, una cada 15 minutos.

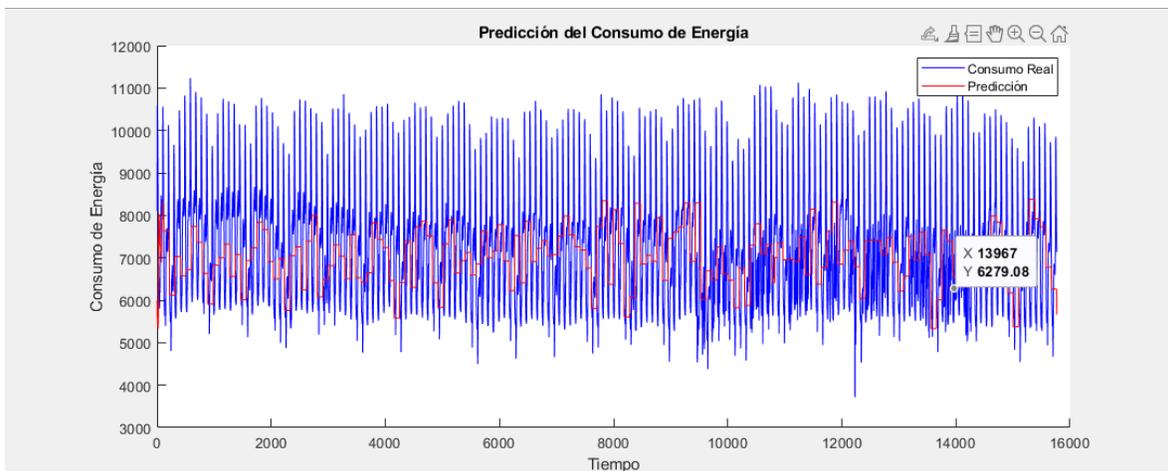


Figura 17.- Predicción Consumo Real vs Consumo Predicho con el modelo SARIMAX (0,1,0(0,0,0)).

La Tabla 8 muestra el Error Porcentual Absoluto Medio (MAPE), Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE).

ERROR MAE	ERROR MAPE	ERROR RMSE
1090.00	14.86%	1373.20

Tabla 8.- MAE, MAPE y RMSE para modelo SARIMAX.

En el anexo 1, se encuentra el registro de desarrollo del código además que el código completo con los comentarios respectivos se encuentran en el repositorio del proyecto.

3.11.2 Desarrollo del modelo híbrido (SARIMAX con Redes Neuronales)

Con base en [10], en este proyecto, se utilizó una Red Neuronal Artificial (RNA) en combinación con un modelo SARIMAX para mejorar la precisión en la predicción del consumo eléctrico. La red neuronal permitió capturar patrones no lineales en los datos, complementando la capacidad del modelo estadístico de identificar tendencias y estacionalidad.

El modelado del componente SARIMAX ha sido previamente especificado en el apartado 3.11.1, donde se detallan su configuración, selección de hiperparámetros y proceso de entrenamiento. En este apartado, se enfoca la integración de SARIMAX con la red neuronal para mejorar la precisión de la predicción del consumo eléctrico.

- **Desarrollo de la Red Neuronal**

a. Arquitectura del modelo

La red neuronal utilizada en este proyecto es un Perceptrón Multicapa (MLP), un tipo de red neuronal artificial que permite modelar relaciones no lineales en los datos, que se entrena con el algoritmo Levenberg-Marquardt (`trainlm`). Su arquitectura se diseñó considerando la estructura de la serie temporal y la influencia de variables exógenas, su objetivo es corregir los residuos del modelo SARIMAX, aprendiendo a predecir los mismos en función de las variables exógenas, luego, genera predicciones de residuos en el conjunto de prueba (`residuals_pred`). A continuación se desarrolla más a fondo los componentes de la arquitectura de la red neuronal con ayuda de la Figura 18.

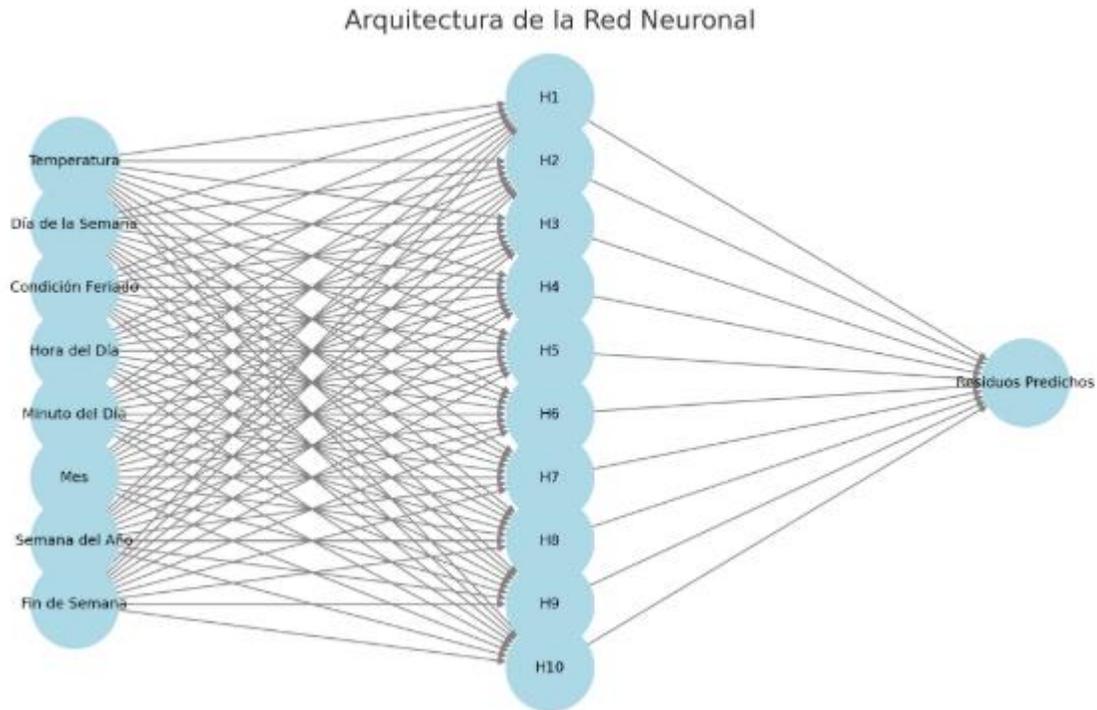


Figura 18.- Arquitectura de la Red Neuronal Recurrente (RNN) utilizada en el modelo híbrido.

1. Capa de Entrada

La red recibe como entrada un conjunto de 8 variables exógenas:

- Temperatura
- Día de la Semana (DiaNum)
- Condición de Feriado (Condición)
- Hora del Día
- Minuto del Día
- Mes
- Semana del Año
- Fin de Semana (1 = Sábado/Domingo, 0 = Día Laboral)

Estas variables son normalizadas mediante la función mapminmax, que ajusta los valores al rango [-1,1] para mejorar la eficiencia del entrenamiento.

2. Capa Oculta

Número de capas ocultas: 1

Número de neuronas en la capa oculta: 10

Función de activación: Por defecto, MATLAB usa la sigmoide logística (logsig) en redes fitnet.

3. Capa de Salida

La red tiene una sola neurona de salida, encargada de predecir los residuos del modelo SARIMAX. La salida es desnormalizada con mapminmax ('reverse', residuals_pred_norm, Ys) para devolverla a su escala original.

b. Entrenamiento del Modelo

- Algoritmo de entrenamiento: trainlm (Levenberg-Marquardt).
- Número de épocas: 100 iteraciones para minimizar el error.
- Criterio de parada: Se detiene el entrenamiento si el gradiente es menor a 1e-6.

- **Integración con SARIMAX**

La predicción final combinó las salidas del modelo SARIMAX con la red neuronal, obteniendo un modelo híbrido que mejora la captura de patrones complejos en el consumo eléctrico. El código se encarga de analizar los residuos del modelo híbrido y comparar los valores predichos con los reales en un día aleatorio de la base de datos de prueba.

- **Implementación en MATLAB**

Se aplican las variables de prueba a la red neuronal entrenada para obtener los residuos predichos.

La predicción final se obtiene al sumar la predicción del modelo SARIMAX con la corrección de la red neuronal:

$$Y_{\text{pred}_{\text{final}}} = Y_{\text{pred}_{\text{sarima}}} + \text{residuals}_{\text{pred}} \quad (2)$$

Este diseño de la red neuronal permite capturar dinámicas complejas del consumo eléctrico, complementando el enfoque estadístico de SARIMAX y mejorando la capacidad del modelo híbrido para adaptarse a variaciones en la demanda energética.

Esta estructura permitió capturar relaciones no lineales en los datos, mejorando la precisión del modelo híbrido SARIMAX + Red Neuronal en la predicción del consumo eléctrico.

- **Predicción y Evaluación del modelo**

Una vez entrenado el modelo, se realizaron predicciones sobre el conjunto de prueba. Se calcularon las métricas de evaluación: Error Porcentual Absoluto Medio (MAPE), Error

Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE), en la Tabla 9, se evidencia el resultado de estos.

MAE	MAPE	RMSE
7216.59	101.91%	7286.80

Tabla 9.- MAE, MAPE y RMSE para el modelo de SARIMAX + RNN.

En la Figura 19 se presentan los resultados de la predicción utilizando SARIMAX+Redes Neuronales, como se observa en la gráfica la predicción resulta en valores muy elevados; se generó una tabla con los resultados cada 15 minutos, donde se evidencia el amplio error que existe, esto acompañado de los errores en la Tabla 9 permite descartar la opción del modelo híbrido para este caso.

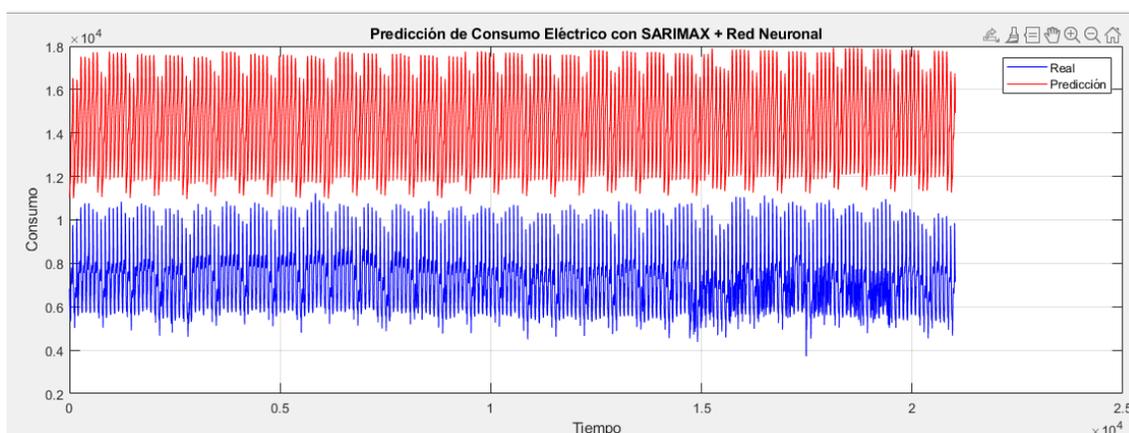


Figura 19.- Comparación del consumo real vs predicción del modelo híbrido.

En el anexo 2, se encuentra el registro de desarrollo del código además que el código completo, con los comentarios respectivos como registro se encuentran en el repositorio del proyecto.

3.11.3 Desarrollo del modelo LSTM (Long Short-Term Memory)

En este estudio, se emplea un modelo LSTM para predecir el consumo eléctrico en función de múltiples variables exógenas, como la temperatura, el día de la semana y la condición de feriado. Debido a su capacidad para capturar patrones temporales complejos y retener información relevante a lo largo del tiempo, las LSTM resultan particularmente adecuadas para esta tarea, permitiendo mejorar la precisión de las predicciones y facilitar una mejor gestión del consumo energético.

a. Arquitectura de la Red Neuronal LSTM

Se diseñó una red neuronal basada en Long Short-Term Memory (LSTM), una variante de redes neuronales recurrentes (RNN) que permite capturar dependencias temporales a largo plazo en series de tiempo. En la Figura 20 se observa la representación de la arquitectura de la red neuronal implementada y las capas que la componen:



Figura 20.- Arquitectura del modelo LSTM

1. Capa de entrada (`sequenceInputLayer`):

Recibe secuencias de datos que incluyen múltiples variables exógenas y el consumo eléctrico histórico. La dimensión de entrada está determinada por el número total de características utilizadas en el modelo, lo que permite a la red aprovechar la información contextual de las variables exógenas junto con la serie temporal del consumo. En este caso, la red toma en cuenta los valores de consumo eléctrico de los últimos cuatro pasos de tiempo ($\text{lag} = 4$), así como factores externos como la temperatura, el día de la semana, la condición de feriado y características derivadas de la fecha y hora (hora del día, minuto del día, mes, semana del año y si es fin de semana).

2. Capa LSTM (`lstmLayer`)

Contiene 50 neuronas ocultas y se configura con el modo de salida `last`, lo que significa que solo se toma en cuenta el último estado oculto de la secuencia para la predicción. Esta capa permite que la red retenga información relevante de los pasos temporales anteriores.

Se configura con el parámetro `OutputMode = 'last'`, lo que significa que la salida de la capa LSTM es el estado oculto correspondiente al último paso de la secuencia. Esta capa es responsable de aprender las relaciones temporales entre los datos y retener información relevante de eventos pasados que influyen en el consumo eléctrico.

3. Capa completamente conectada (`fullyConnectedLayer`):

Reduce la dimensionalidad de la salida de la LSTM a un solo valor escalar, correspondiente a la predicción del consumo eléctrico en el siguiente instante de tiempo. Actúa como un transformador que convierte la salida de la LSTM en una estimación concreta de consumo.

4. Capa de regresión (regressionLayer)

Se emplea para calcular la función de pérdida del modelo durante el entrenamiento. En este caso, la métrica utilizada es el error cuadrático medio (MSE), que mide la diferencia entre las predicciones y los valores reales. A partir de esta métrica, el modelo ajusta sus pesos mediante la retropropagación para minimizar la diferencia entre las predicciones y los valores reales de consumo.

b. Configuración del Entrenamiento

El modelo se entrenó utilizando el algoritmo de optimización Adam con los siguientes parámetros:

- **Número de épocas:** 200, lo que define el número total de iteraciones sobre el conjunto de entrenamiento.
- **Tasa de aprendizaje inicial:** 0.005, con un esquema de reducción piecewise que ajusta la tasa de aprendizaje en la época 125 con un factor de 0.2.
- **Umbral de gradiente:** 1, para evitar explosión de gradientes y mejorar la estabilidad del entrenamiento.
- **Modo de verbosidad:** Desactivado para evitar exceso de salida en consola.

El conjunto de entrenamiento representó el 60% de los datos disponibles, mientras que el 40% restante se utilizó para la validación y prueba del modelo.

c. Normalización de Datos

Para mejorar la estabilidad del entrenamiento, se normalizaron tanto las secuencias de entrada como las salidas. Se calcularon la media y desviación estándar a partir del conjunto de entrenamiento y se aplicaron funciones de normalización:

d. Predicción y Evaluación del Modelo

Una vez entrenado el modelo, se realizaron predicciones sobre el conjunto de prueba. Las predicciones normalizadas fueron desnormalizadas para comparar con los valores reales. Se calcularon las métricas de evaluación: Error Porcentual Absoluto Medio (MAPE), Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE), en la Tabla 10 se evidencia el resultado de estos.

MAPE	MAE	RMSE
0.90%	63.50	130.92

Tabla 10.- MAPE, MAE y RMSE para el modelo LSTM.

La Figura 21 indica los resultados obtenidos del modelo LSTM para la predicción de consumo de energía eléctrica. En el anexo 3, se encuentra el registro de desarrollo del código además que el código completo, con los comentarios respectivos como registro se encuentran en el repositorio del proyecto.

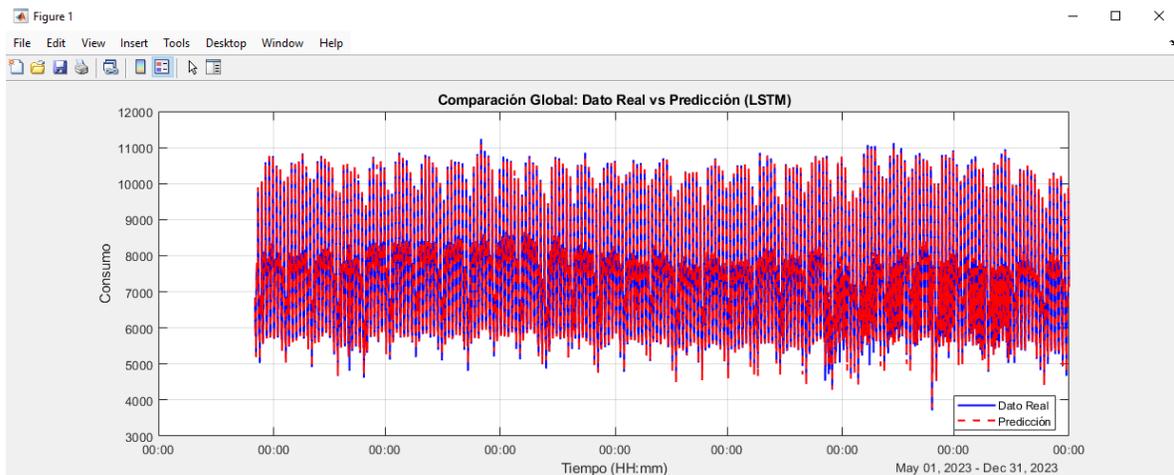


Figura 21.- Comparación del Consumo Real vs la predicción utilizando LSTM.

3.12 VALIDEZ DEL MODELO MEJOR AJUSTADO

Una vez que se han calculado los errores de los modelos (MAE, RMSE y MAPE), se tiene una idea preliminar de cuál es el modelo que se ajusta mejor a los datos en términos de precisión de la predicción. Sin embargo, esos errores por sí solos no son suficientes para determinar si un modelo es aceptable en el contexto. Por esta razón, es necesaria la interpretación de los errores, para saber si es aceptable en el contexto.

Un error más bajo no siempre significa o implica que un modelo sea válido, con base en los errores, los modelos son aceptables, para este contexto, si: el valor del RMSE y MAE es bajo en proporción al valor medio del consumo y si el MAPE y el RMSE (%) no supera el 10% (MAPE<10%).

En la Tabla 11 se observa el resumen de la interpretación de los diferentes errores para los modelos en cuestión.

	LSTM	SARIMAX	SARIMAX+RNN
MAE vs Valor medio de Consumo (kWh)	63.50 << 7204,83	1090.00 < 7204,83	7216.59 > 7204,83
MAPE vs MAPE aceptable (10%)	0.90% << 10%	14.86% > 10%	101.91% >> 10%
RMSE (kWh) vs Valor medio de Consumo	130.92 << 7204,83	1373.20 < 7204,83	7286.80 > 7204,83
RMSE (%) vs RMSE aceptable (10%)	1.8 % << 10%	19.1% > 10%	101.13% >> 10%

Tabla 11.- Interpretación de los errores

Con base en la Tabla 11, se determina que el modelo LSTM se ajusta a la naturaleza de los datos y es aceptable en el contexto de consumo de energía eléctrica. El valor del MAPE indica que los datos predichos tienen un error menor al $\pm 1\%$. Con base en el valor RMSE porcentual de 1,8% se determina un excelente ajuste, mucho menor que el 10% que es el error válido al hablar en términos de consumo de energía. Los valores MAE y RMSE son bastante subjetivos puesto que se determina su validez dependiendo de qué tan altos son en comparación a la media del sistema, sin embargo para el modelo LSTM los valores son mucho más pequeños que la media. Sin embargo, esta validez tiene ciertos límites y no garantiza que el modelo sea perfecto o que no pueda ser mejorado.

CAPÍTULO IV. RESULTADOS Y DISCUSIÓN.

4.1. ESTACIONALIDAD DEL CONSUMO ELÉCTRICO

El análisis de los datos de consumo eléctrico en intervalos de 15 minutos para diferentes días del año revela un patrón de estacionalidad diaria bien definido, en el intervalo de tiempo entre las 00:00 - 06:00, se registra un consumo relativamente bajo, con una ligera disminución progresiva hasta alcanzar el punto más bajo alrededor de las 04:00 - 05:00. A partir de las primeras horas del día, se observa un incremento en el consumo, con un aumento significativo entre las 06:00 y las 08:00, lo que puede atribuirse al inicio de actividades laborales y domésticas. Durante el mediodía y primeras horas de la tarde, el consumo se estabiliza en un nivel intermedio, hasta la noche (18:00 - 00:00), donde se evidencia un pico de consumo entre las 18:00 y las 21:00, el cual coincide con el retorno de actividades laborales y el uso de iluminación, electrodomésticos y otros equipos eléctricos en los hogares. Posteriormente, el consumo comienza a descender de manera gradual. Estos patrones se repiten de manera consistente en diferentes días, lo que confirma la existencia de una estacionalidad diaria marcada en el consumo eléctrico.

4.2 ANÁLISIS DE VARIABLES EXÓGENAS

Se realizó un análisis censal para determinar la relación entre las variables exógenas de la temperatura, el día de la semana y la condición de feriado con el consumo eléctrico. Los resultados indicaron que:

- La Temperatura Media mostró tener influencia sobre el consumo de energía eléctrica en la ciudad, por lo que se considera como una variable exógena en el estudio.
- Se identifica diferencias en el consumo eléctrico entre los días laborables y los fines de semana, por lo que se considera una variable exógena en el sistema.
- Se determinó que la condición de feriado influye en el consumo energético, generando variaciones significativas respecto a los días regulares, por lo que se determina que también es una variable exógena.
- Los días de fin de semana tienden a presentar un consumo menor en comparación con los días laborables, lo que sugiere una menor actividad industrial y comercial en estos días.
- Las variables de día de la semana y condición de feriado presentaron patrones estacionales claros, lo que sugiere que la demanda energética sigue un comportamiento cíclico en función de estos factores.

4.3 COMPARACIÓN DE MODELOS

Para evaluar el desempeño de los modelos en la predicción del consumo eléctrico, se implementaron y compararon tres enfoques distintos: SARIMAX, SARIMAX combinado con redes neuronales y LSTM. La comparación se realizó utilizando tres métricas de error ampliamente aceptadas en series de tiempo: MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) y RMSE (Root Mean Squared Error).

La Tabla 12 resume los errores obtenidos para cada modelo:

Modelo	SARIMAX	SARIMAX + RNN	LSTM
Error MAE	1090.00	7216.59	63.50
Error MAPE	14.86%	101.91%	0.90%
Error RMSE	1373.20	7286.80	130.92

Tabla 12.-Comparación de errores MAE, MAPE y RMSE para los modelos planteados.

Los resultados muestran que el modelo LSTM obtuvo el menor error en todas las métricas evaluadas, lo que indica una mejor capacidad para capturar la dinámica temporal del consumo eléctrico. Esto se debe a la capacidad de las redes LSTM para manejar dependencias a largo plazo y patrones complejos en series de tiempo.

El modelo SARIMAX, si bien es útil para capturar patrones estacionales y tendencias en la serie, presentó mayores errores en comparación con los otros enfoques, lo que sugiere que no logra modelar con precisión las fluctuaciones de corto plazo en el consumo eléctrico. Por otro lado, la combinación de SARIMAX con redes neuronales mejoró ligeramente el rendimiento sin embargo generaba valores muy altos lograba capturar el patrón pero no los valores, aun cuando existió normalización de datos, por lo que no logró superar el desempeño del modelo basado únicamente en LSTM.

4.4 EVALUACIÓN DEL MODELO LSTM

Dado que el modelo LSTM mostró el mejor desempeño, se realizó un análisis más detallado de sus predicciones. En la Figura 22 se presenta una comparación entre los valores reales y los valores predichos por el modelo.

Se observó que:

- El modelo capturó con precisión las tendencias generales del consumo eléctrico.
- A pesar de algunos errores en los picos de consumo, el modelo mantiene una estabilidad adecuada en su desempeño.

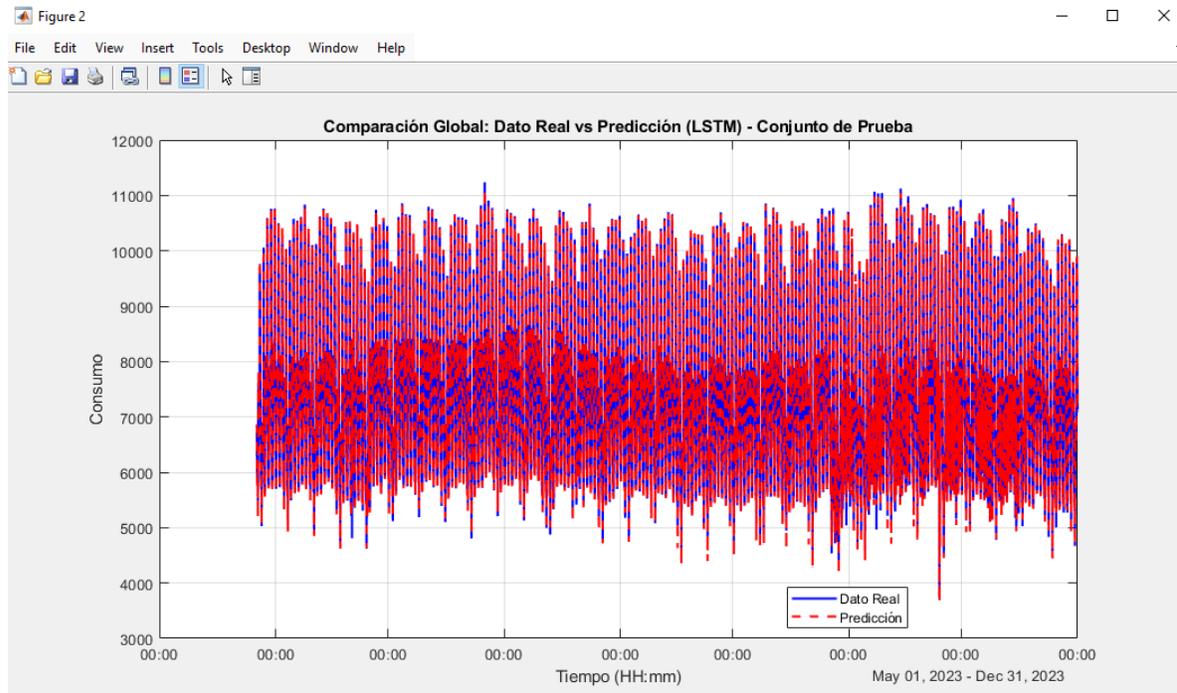


Figura 22.- Comparación entre los consumos reales y la predicción utilizando el modelo seleccionado.

En la Figura 23 se observa la evolución del consumo energético en intervalos de 15 minutos a lo largo del día, donde la línea azul representa el consumo real y la línea roja discontinua indica el consumo predicho por el modelo LSTM. Se evidencia que el modelo logra capturar con alta precisión la tendencia del consumo eléctrico, siguiendo de cerca las variaciones a lo largo del día. Las principales discrepancias se presentan en momentos de cambios bruscos en la demanda, como el pico en horas de la mañana.

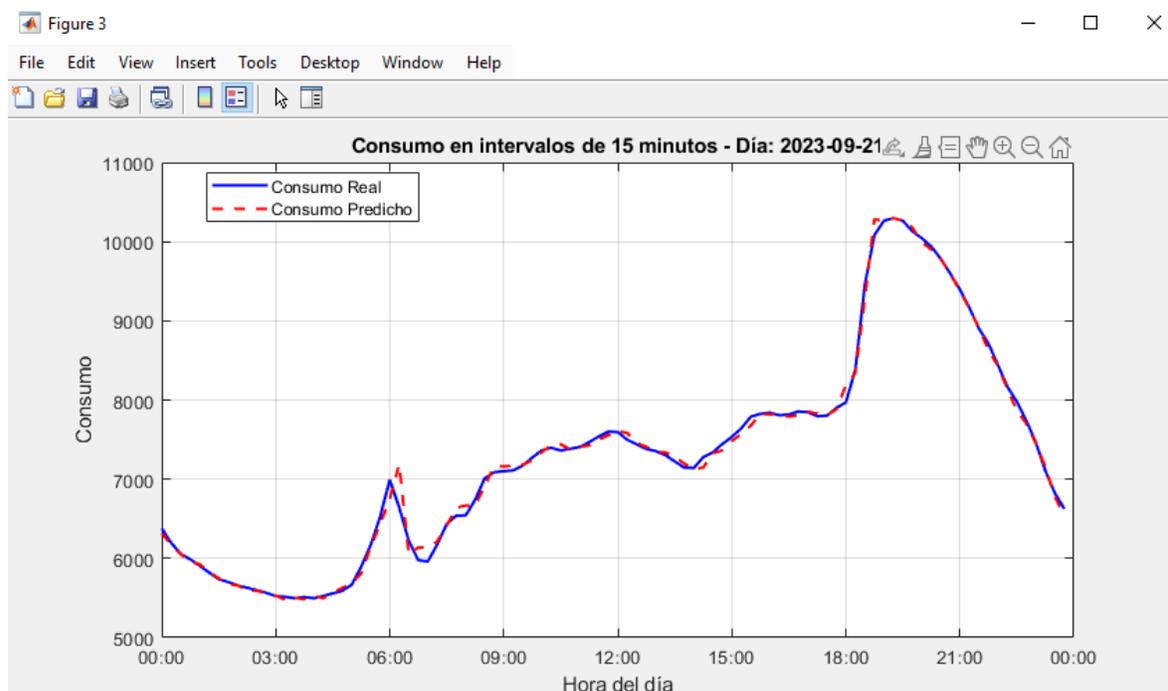


Figura 23.- Comparación entre el consumo real y el consumo predicho por el modelo LSTM para el 21 de septiembre de 2023.

Finalmente, para facilitar la visualización y el uso de los datos, el sistema genera una tabla con las predicciones en intervalos de 15 minutos. No obstante, debido a la extensión de esta información, no es posible presentarla en su totalidad en el siguiente apartado. Por ello, se ha resumido en una tabla que muestra los resultados mensuales de la predicción junto con los datos reales. Los datos completos se pueden encontrar en el repositorio del proyecto al que se puede acceder en el siguiente enlace: <https://github.com/Vpvicpar/Codigos-Predictivos-EERSA>, en el archivo resultados_datos_LSTM.

Mes	Dato Real	Predicción
jun-23	21200724,38	21165277,4
jul-23	22336807,66	22284621,6
ago-23	22046822,46	22020357,8
sep-23	20770229,7	20760033,5
oct-23	21377224	21357827,1
nov-23	20692702,95	20656567,7
dic-23	20435772,28	20416130

Tabla 13.- Resultados de la predicción realizada con el modelo LSTM.

4.5 VALIDACIÓN DEL MODELO

Para validar la confiabilidad del modelo LSTM, se utilizó un conjunto de datos de prueba independiente, extraído del conjunto total de datos disponibles correspondiente al 20% de este. Se encontró que los errores se mantuvieron dentro de un rango aceptable, lo que indica que el modelo no presenta problemas de sobreajuste y puede generalizarse adecuadamente a datos no vistos.

El propósito de la validación es evaluar no solo la precisión de las predicciones, sino también determinar si el modelo tiene un error aceptable y si es válido en el contexto. El valor del MAPE indica que los datos predichos tienen un error menor al $\pm 1\%$. Con base en el valor RMSE porcentual de 1,8% se determina un excelente ajuste, mucho menor que el 10% que es el error válido al hablar en términos de consumo de energía. Los valores MAE y RMSE son bastante subjetivos puesto que se determina su validez dependiendo de qué tan altos son en comparación a la media del sistema, sin embargo para el modelo LSTM los valores son mucho más pequeños que la media. Sin embargo, esta validez tiene ciertos límites y no garantiza que el modelo sea perfecto o que no pueda ser mejorado.

4.6 IMPLICACIONES Y APLICABILIDAD

Los resultados obtenidos en este estudio sugieren que el modelo LSTM es una herramienta viable para la predicción del consumo eléctrico en la ciudad de Riobamba. Su capacidad para capturar patrones temporales complejos lo hace útil para:

- Mejorar la planificación del suministro energético en la ciudad, permitiendo a las empresas distribuidoras de energía anticipar la demanda y ajustar la generación de manera eficiente.
- Apoyar la toma de decisiones en políticas de eficiencia energética, facilitando la implementación de estrategias que promuevan el uso responsable de la electricidad.
- Reducir costos asociados a la compra y distribución de energía, al mejorar la precisión en la estimación de la demanda y minimizar pérdidas.
- Facilitar la integración de fuentes de energía renovable en la matriz energética local, al proporcionar predicciones más precisas que permitan una mejor gestión de recursos intermitentes como la energía solar o eólica.
- Desarrollar estrategias de respuesta a la demanda, donde los consumidores puedan recibir incentivos para reducir su consumo en momentos de alta demanda.

En futuras investigaciones, se recomienda explorar arquitecturas más avanzadas de redes neuronales y la inclusión de nuevas variables exógenas que puedan mejorar la precisión del modelo. Además, la implementación de técnicas de optimización y ajuste de hiperparámetros podría contribuir a un desempeño aún más robusto del sistema de predicción.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.

5.1 CONCLUSIONES

- El análisis de los datos de consumo eléctrico en Riobamba durante los años 2021, 2022 y 2023 permitió identificar patrones de estacionalidad diaria y semanal. Se evidenció que el consumo sigue un comportamiento cíclico, con picos de demanda entre las 18:00 y las 21:00 horas. Además, el análisis estadístico confirmó que las variables de la temperatura media, el día de la semana y la condición de feriado influyen en el consumo eléctrico. Se determinó que los fines de semana presentan menor demanda energética en comparación con los días laborables, y que los feriados generan variaciones en el consumo.
- Se evaluaron tres modelos para la predicción del consumo eléctrico: SARIMAX, SARIMAX combinado con redes neuronales y LSTM. Los resultados demostraron que el modelo basado en redes neuronales LSTM presentó el mejor desempeño en términos de error y ajuste, con valores significativamente menores en comparación con los otros enfoques. Esto confirma que las redes LSTM tienen una mejor capacidad para capturar patrones temporales complejos en series de tiempo, especialmente en la predicción de datos con variaciones dinámicas como el consumo energético.
- La validación del modelo LSTM mediante un conjunto de datos de prueba independiente confirmó que sus predicciones mantienen un nivel de error dentro de un rango aceptable, sin evidencias de sobreajuste.
- En el contexto ecuatoriano, donde la matriz energética depende en gran medida de fuentes hidroeléctricas y está sujeta a variaciones climáticas (como sequías que afectan la generación), es importante mejorar la eficiencia en el consumo y en la planificación de la distribución. A partir de los hallazgos del estudio, se sugieren las siguientes estrategias:
 - a) Utilizar las predicciones para mejorar la operación de las centrales hidroeléctricas y térmicas, reduciendo la necesidad de generación de emergencia en momentos de alta demanda.
 - b) Programar mantenimientos de infraestructura eléctrica en períodos de baja demanda para minimizar el impacto en los usuarios.
 - c) Aplicar incentivos económicos a usuarios residenciales, comerciales e industriales que trasladen su consumo a horarios de menor demanda, como se ha hecho en otros países con esquemas de "tarifas horarias inteligentes".
 - d) Implementar sistemas de iluminación LED con sensores de ocupación en las calles y espacios públicos para reducir el consumo nocturno innecesario.
 - e) Promover campañas dirigidas a la ciudadanía sobre hábitos de consumo responsable, priorizando la reducción del consumo en horas pico.

- f) Fomentar la instalación de paneles solares fotovoltaicos en viviendas, industrias y edificios públicos, incentivando la autogeneración con esquemas de "net metering" para devolver el excedente de energía a la red.
- g) Utilizar los modelos de predicción para mejorar la planificación de parques eólicos y solares, ajustando su producción según la demanda proyectada.
- h) Evaluar la viabilidad de sistemas de almacenamiento con baterías para estabilizar la red y reducir la dependencia de fuentes fósiles en momentos de baja generación hidroeléctrica.
- i) Promover el uso de medidores eléctricos inteligentes que permitan a los consumidores monitorear y ajustar su consumo en tiempo real.

5.2. RECOMENDACIONES

- Se recomienda fortalecer el monitoreo del consumo energético mediante la recopilación y análisis continuo de datos en tiempo real; implementando un sistema de monitoreo automatizado que registre el consumo eléctrico en la ciudad y analice su evolución en función de variables externas; desarrollando reportes periódicos que permitan a las autoridades y empresas eléctricas anticipar cambios en la demanda y tomar decisiones basadas en datos e incorporando nuevos factores en los estudios futuros, como el impacto de eventos climáticos extremos y cambios en el crecimiento urbano.
- Para mejorar aún más la precisión del sistema predictivo, se recomienda estudios futuros del modelo donde se incorporen nuevas variables como humedad, radiación solar y nivel de actividad económica en la ciudad así como también explorar el uso de modelos híbridos, combinando redes neuronales con técnicas estadísticas como XGBoost, para capturar mejor las fluctuaciones del consumo.
- Se recomienda la implementación de un sistema de gestión de la demanda basado en predicciones de consumo energético, utilizando el modelo LSTM desarrollado en este estudio, pero también que se incluya elementos de medición constante en los medidores de cada consumidor de manera que la base de datos permita expandir la funcionalidad de este proyecto a consumo por sectores (residencial, comercial o industrial) o por zonas (parroquias, barrios o incluso determinadas calles) y que al manejar datos más exactos, el modelo reduzca el error al mínimo posible.
- Aprovechar la capacidad del modelo y cruzarlo con datos de generación de energía renovable (solar, eólica) para mejorar la integración de estas fuentes en la matriz energética del país, diseñar estrategias específicas para cada sector, promover el uso de energía renovables entre la población que menos consume y concientizar a los sectores que más consumen y permitir que las empresas generadoras y proveedoras continúen siendo públicas pero a su vez que puedan tener planes de compra de energía sin desperdiciar energía o a su vez para que no se sufra de desabastecimiento.
- Se recomienda la ampliación del estudio mediante la incorporación de datos de energía reactiva en el modelo de predicción del consumo energético. Esto permitiría desarrollar un sistema más integral que no solo anticipe la demanda de energía activa, sino que también optimice el factor de potencia en tiempo real, se sugiere la integración de un sistema automatizado de gestión energética, que utilice las predicciones generadas para activar mecanismos de compensación reactiva, como bancos de capacitores o dispositivos FACTS, optimizando así la eficiencia energética y reduciendo penalizaciones por bajo factor de potencia.

BIBLIOGRAFÍA

- [1] Ministerio de Energía y Minas, «Comunicado Oficial: Emergencia Sector Eléctrico- Ministerio de Energía y Minas,» 18 octubre 2023. [En línea]. Available: <https://www.recursoyenergia.gob.ec/comunicado-oficial-emergencia-sector-electrico/>.
- [2] Ministerio de Energía y Minas, «Gobierno Nacional propone alternativas para hacer frente a la emergencia que atraviesa el sector eléctrico,» 25 noviembre 2023. [En línea]. Available: <https://www.recursoyenergia.gob.ec/gobierno-nacional-propone-alternativas-para-hacer-frente-a-la-emergencia-que-atraviesa-el-sector-electrico/>.
- [3] Redacción Primicias, «USD 400 millones le cuesta al país resolver problema de estiaje,» *PRIMICIAS*, 4 enero 2023.
- [4] La Hora, «Colombia se comprometió a vender electricidad más barata a Ecuador,» *Diario La Hora*, 14 diciembre 2023.
- [5] Ministerio de Energía y Recursos Naturales no Renovables, «Plan Maestro de Electricidad,» 2018.
- [6] P. Ullán, R. Gonzalez y M. Gonzalez, *Redes LSTM para la predicción de series temporales.*, Salamanca: Universidad de Salamanca, 2021.
- [7] S. Abdul Khaliq, N. Amril, K. Nadia, S. S. Abdul y A. Naresh, «A new approach to seasonal energy consumption forecasting using temporal convolutional networks,» *Results in Engineering*, vol. 19, p. 101296, septiembre 2023.
- [8] A. Saranj y M. Zolfaghari, «The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models,» *Energy Reports*, vol. 8, pp. 7657-7679, noviembre 2022.

- [9] T. Le, M. Thanh Vo, B. Vo, E. Hwang, S. Rho y S. Wook Baik, «Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM,» *Applied Sciences*, vol. 9, nº 20, 10 octubre 2019.
- [10] L. Ortega Diaz, J. Cárdenas Rangel y G. Osma Pinto, «Estrategias de predicción de consumo energético en edificaciones: una revisión,» *TecnoLógicas*, vol. 26, nº 58, p. 34, 11 Septiembre 2023.
- [11] E. Mejía y S. Gonzales, «Predicción del consumo de energía eléctrica residencial de la Región Cajamarca mediante modelos Holt -Winters,» *Energética*, vol. 40, nº 3, 18 septiembre 2019.
- [12] A. Manobanda, *Predicción de la demanda de energía eléctrica en la producción de petróleo de los campos de Petroamazonas Ep utilizando redes neuronales artificiales.*, Quito: EPN, 2020, p. 114.
- [13] J. Balcells, J. Autonell, V. Barra, J. Brossa, F. Fornieles, B. Garcia, J. Ros y J. Serra, *Eficiencia en el uso de la Energía Eléctrica*, Bogotá: Alfaomega Marcombo, 2016, p. 336.
- [14] Konrad-Adenauer-Stiftung, «Gestión de Energía Eléctrica,» 2021. [En línea]. Available: <https://igecc-mex.org/wp-content/uploads/2021/04/Gestio%CC%81n-de-energi%CC%81a-ele%CC%81ctrica.pdf>.
- [15] EERSA, «Empresa Eléctrica Riobamba S.A.,» [En línea]. Available: <https://www.eersa.com.ec/site/>. [Último acceso: 2024].
- [16] EERSA, «Geoportal Eléctrico EERSA,» 2024. [En línea]. Available: <https://portal4arcgis.eersa.com.ec/portal/apps/webappviewer/index.html?id=c7bfb17dd21248bbbea20af211b8388e>. [Último acceso: 17 Agosto 2024].

- [1 B. Mahesh, «Machine Learning Algorithms - A Review,» *International Journal of Science and Research (IJSR)*, vol. 9, nº 1, enero 2020.
- [1 L. Sandoval, «Algoritmos de aprendizaje automático para análisis y predicción de datos,»
8] *Revista Tecnológica*, nº 11, 2018.
- [1 C. Miranda, *Modelización de Series Temporales modelos clásicos y SARIMA*, Granada,
9] 2021.
- [2 C. Arana, *Redes neuronales recurrentes: Análisis de los modelos especializados en datos
0] secuenciales*, Buenos Aires: UCEMA, 2021.
- [2 D. A. Restrepo Leal, J. P. Vilorio Porto y C. A. Robles Algarín, El camino a las redes
1] neuronales artificiales, vol. 1, Santa Marta: Unimagdalena, 2021, p. 204.
- [2 Z. Moslemi, L. Clark, S. Kernal, S. Rehome, S. Sprengel, A. Tamizifar, S. Tuli, V.
2] Chokshi, M. Nomeli, E. Liang, M. Bidgol, J. Lu, M. Dasaur y M. Hodgett,
«Comprehensive Forecasting of California's Energy Consumption: A Multi-Source and
Sectoral Analysis Using ARIMA and ARIMAX Models,» *arXiv preprint
arXiv:2402.04432*, Agosto 2023.
- [2 Z. Yagang, H. Jingyi, P. Guifang, X. Yan y W. Fei, «A multi-stage predicting
3] methodology based on data decomposition and error correction for ultra-short-term wind
energy prediction,» *Journal of Cleaner Production*, vol. 292, nº 125981, 2021.
- [2 X. Genglei, B. Yuepeng y W. Chenyang, «Optimization design of passive residual heat
4] removal system based on improved genetic algorithm,» *Annals of Nuclear Energy*, vol.
189, nº 109859, 2023.

ANEXOS

1. ANEXO 1. CARTA DE COMPROMISO



Empresa Eléctrica Riobamba S.A.

eersa

Riobamba, 16 de febrero de 2024

A quien corresponda:

Por medio de la presente, la Empresa Eléctrica Riobamba S.A. se compromete formalmente a entregar los datos sobre el consumo de energía eléctrica registrados por los medidores en los años 2021, 2022 y 2023, en cumplimiento con los acuerdos establecidos y en apego a las normativas de confidencialidad y protección de datos aplicables.

Los datos proporcionados incluirán las mediciones registradas en los periodos especificados, asegurando su integridad y veracidad, con el objetivo de apoyar el análisis y estudio correspondiente. Asimismo, se garantizará que la información se entregue en el formato acordado y dentro del plazo establecido entre ambas partes.

Los datos entregados serán entregados exclusivamente a Victoria Estefanía Parra Rocha, estudiante de la Universidad Nacional de Chimborazo con documento de identidad 0604537779, para su utilización en el proyecto de investigación titulado "Desarrollo de un sistema predictivo de consumo de energía eléctrica en la ciudad de Riobamba mediante la aplicación de algoritmos de predicción para la mejora de la gestión energética".

EERSA declara que la información entregada será utilizada exclusivamente para los fines previamente acordados y bajo los términos de confidencialidad correspondientes.

Sin otro particular, quedamos a su disposición para cualquier aclaración adicional.

Atentamente,

Ing. Augusto Benjamín Guerrero Lara
0602776866

GERENTE
Empresa Eléctrica Riobamba S.A.

Trabajamos para iluminar tu vida ...

2. ANEXO 2. REGISTRO DE DESARROLLO DEL MODELO SARIMAX

Registro de Desarrollo			
Nombre: Victoria Parra Rocha		Modelo: SARIMAX	
Observaciones: El código principal se puede encontrar en el repositorio del proyecto, donde se agregan comentarios que permitieron llevar un registro detallado del mismo y permiten manipular variables e hiperparámetros.			
Fecha	Código	Errores	Soluciones
Agosto-2024	datos_consumo_energia = datos_consumo_energia.Dia	Las variables exógenas de día de la semana y feriado son de tipo texto.	Convertir las variables exógenas a variables tipo dummy datos_consumo_energia = [datos_consumo_energia, ... dummyvar(categorical(datos_consumo_energia.Dia))];
Agosto 2024	datos_consumo_energia = [datos_consumo_energia, ... dummyvar(categorical(datos_consumo_energia.Dia))];	Error using tabular/horzcat (line 223) All input arguments must be tables.	Crear tablas convirtiendo las variables de texto en variables numéricas. Los días de la semana son números enteros del 1 al 7 y los feriados se denominan como 1 y los días normales con 0.
Agosto 2024	mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12, 'Exogenous', size(XEntrenamiento, 2));	Error using Prediccion2 (line 29) 'Exogenous' is not a recognized parameter. For a list of valid name-value pair arguments, see the documentation for this function.	% Definir el modelo ARIMA con estacionalidad para incluir las variables exógenas después mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12); % Estimar el modelo ARIMA con los predictores exógenos modelo = estimate(mdl, yEntrenamiento, 'X',

			XEntrenamiento);
Septiembre 2024	mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12);	Error using arima/estimate (line 819) Number of observations of predictors must be at least 73592. Error in Prediccion2 (line 31) modelo = estimate(mdl, yEntrenamiento, 'X', XEntrenamiento);	Revisar la cantidad de datos de entrada y la cantidad de datos que se están generando: Asegurarse que XEntrenamiento tenga el mismo número de filas que la variable dependiente yEntrenamiento: % Verificar las dimensiones de los conjuntos de datos disp('Tamaño de los datos de entrenamiento:'); disp(size(datosEntrenamiento)); disp('Tamaño de los datos de validación:'); disp(size(datosValidacion)); disp('Tamaño de los datos de prueba:'); disp(size(datosPrueba));
Septiembre 2024	mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12);	Error using arima/estimate (line 819) Number of observations of predictors must be at least 73592. Error in Prediccion2 (line 31) modelo = estimate(mdl, yEntrenamiento, 'X', XEntrenamiento);	El error persiste por lo que se busca solucionarlo aumentando el número de observaciones del conjunto de entrenamiento. % Dividir los datos en conjuntos de entrenamiento, validación y prueba totalDatos = size(datos_consumo_energia, 1); tamanoEntrenamiento = round(0.8 * totalDatos); tamanoValidacion = round(0.10 * totalDatos); tamanoPrueba = totalDatos - tamanoEntrenamiento - tamanoValidacion;
Septiembre 2024	mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12);	Error using arima/estimate (line 819) Number of observations of predictors	Se calcula el número mínimo de observaciones necesarias % Número de predictores exógenos (en este caso, se

		<p>must be at least 73592. Error in Prediccion2 (line 31) modelo = estimate mdl, yEntrenamiento, 'X', XEntrenamiento);</p>	<p>puede ajustar según los datos usados) num_predictores_exogenos = 3; % Ejemplo, supongamos que hay 3 variables exógenas (temperatura, día de la semana, feriado) % Calcular el número mínimo de observaciones necesarias observaciones_minimas = max(p, q, P, Q) + s * D + num_predictores_exogenos;</p>
Septiembre 2024	mdl = arima('Constant', 0, 'D', 1, 'Seasonality', 7, 'MALags', 1, 'SMALags', 12);	<p>Error using arima/estimate (line 819) Number of observations of predictors must be at least 73592. Error in Prediccion2 (line 31) modelo = estimate mdl, yEntrenamiento, 'X', XEntrenamiento);</p>	<p>El error persiste por sobre las correcciones que parecen lógicas por lo que se pretende abordar el error desde la generalidad y se parte del modelo ARIMA más simple nuevamente mdl = arima('Constant', 0, 'D', 1, 'MALags', 0,);</p>
Octubre 2024	mdl = arima('Constant', 0, 'D', 1, 'MALags', 0,);	<p>Error using max Dimension argument must be a positive integer scalar, a vector of unique positive integers, or 'all'.</p>	<p>Se cambia el valor de constante por ARLags y los valores que son 0 por positivos enteros mdl = arima('ARLags', 1, 'D', 1, 'MALags', 1, 'SMALags', 96);</p>
Octubre 2024	mdl = arima('ARLags', 1, 'D', 1, 'MALags', 1, 'SMALags', 96);	<p>Error using arima/estimate (line 819) Number of observations of predictors must be at least 73592.</p>	<p>Los hiperparámetros son modificados puesto que se ha determinado con base en teoría que el error se da porque MATLAB necesita mayor cantidad de datos para modelar algoritmos más complejos. Los</p>

		Error in Prediccion2 (line 31) modelo = estimate mdl, yEntrenamiento, 'X', XEntrenamiento);	hiperparámetros con la mejor adaptación a la serie de tiempo y que permiten a MATLAB modelar son: mdl = arima('ARLags', [], 'D', 1, 'MALags', 0, 'SMALags', 96);
--	--	---	--

3. ANEXO 3. REGISTRO DE DESARROLLO DEL MODELO SARIMAX+RNN

Registro de Desarrollo			
Nombre: Victoria Parra Rocha		Modelo: SARIMAX+RNN	
Observaciones: El código principal se puede encontrar en el repositorio del proyecto, donde se agregan comentarios que permitieron llevar un registro detallado del mismo y permiten manipular variables e hiperparámetros. El registro a continuación describe el proceso de adaptación del modelo híbrido haciendo uso del modelo SARIMAX mejor adaptado.			
Fecha	Errores	Observación	Soluciones
Octubre 2024	Dimensions of matrices being concatenated are not consistent.	Inconsistencia en las dimensiones al alimentar la red con datos de entrada y salida.	Se revisaron las dimensiones de los datos asegurando que X_train y Y_train tengan la misma cantidad de observaciones.
Octubre 2024	Error using trainNetwork: Invalid input data type.	La red neuronal esperaba datos en formato cell array o dlarray, pero recibió double.	Se transformaron los datos a cell array usando {} y se usó dlarray para Deep Learning Toolbox.
Octubre 2024	Gradient explosion	Los gradientes de la red explotaban durante el	Se aplicó normalización de datos.

		entrenamiento.	
Noviembre 2024	Slow convergence	El modelo tardaba muchas épocas en mejorar.	Se ajustó la tasa de aprendizaje y se aplicó batch normalization.
Noviembre 2024	Overfitting en datos de entrenamiento.	La red tenía demasiados parámetros para la cantidad de datos.	Se agregó dropout y se redujo la complejidad de la arquitectura.
Noviembre 2024	Predicciones muy cercanas a la media de los datos.	La red no capturaba la estructura temporal correctamente.	Se aumentó el número de lags en los datos de entrada.
Noviembre 2024	Error: Sequence length too short for LSTM.	Se usaron secuencias demasiado cortas para la red.	Se aumentó el window size en la preparación de los datos.
Noviembre 2024	Invalid function call: Missing or extra input arguments.	Problemas al definir la arquitectura de la red en MATLAB.	Se verificó que sequenceInputLayer, lstmLayer, y fullyConnectedLayer estén correctamente configurados.
Noviembre 2024	Inconsistent sample sizes between predictors and target variable.	Desfase entre X_train y Y_train.	Se ajustó la ventana de tiempo en la preparación de datos.
Noviembre 2024	Out of memory error during training.	La red era demasiado grande para la RAM disponible y era innecesario.	Se redujo el tamaño del batch y se utilizó GPU si estaba disponible.
Noviembre 2024	Error using trainNetwork: Inputs must have consistent time steps.	Los datos de entrada a la RNN no tenían el mismo tamaño en todas las	Se usó padsequence() para igualar la longitud de las secuencias.

		muestras.	
Noviembre 2024	Unexpected NaN values in loss function.	Alguna transformación de datos generaba valores no numéricos.	Se verificó que no haya NaN en X_train y Y_train usando isnan().

4. ANEXO 4. REGISTRO DE DESARROLLO DEL MODELO LSTM

Registro de Desarrollo			
Nombre: Victoria Parra Rocha		Modelo: LSTM	
Observaciones: El código principal se puede encontrar en el repositorio del proyecto, donde se agregan comentarios que permitieron llevar un registro detallado del mismo y permiten manipular variables e hiperparámetros.			
Fecha	Errores	Observación	Soluciones
Diciembre 2024	Dimensions of matrices being concatenated are not consistent.	Inconsistencia en las dimensiones de los datos de entrada y salida.	Se verificaron las dimensiones de X_train y Y_train, asegurando que las secuencias sean del mismo tamaño.
Diciembre 2024	Error using trainNetwork: Invalid input data type.	La red esperaba un formato cell array o dlarray, pero recibió double.	Se convirtió X_train a cell array con {} o dlarray.
Diciembre 2024	Gradient explosion	Valores muy altos en los gradientes, causando inestabilidad en el entrenamiento.	Se aplicó gradiente clipping (GradientThreshold) y normalización de los datos.
Diciembre	Slow convergence	El modelo tardaba muchas	Se ajustó la tasa de aprendizaje (adam con

2024		épocas en mejorar.	LearningRateSchedule).
Diciembre 2024	Overfitting en datos de entrenamiento.	La red tenía demasiados parámetros para la cantidad de datos.	Se aplicó dropout, regularización L2 y reducción de unidades LSTM.
Diciembre 2024	Predicciones constantes o sin variación.	El modelo no aprendía patrones temporales.	Se aumentó el número de lags en los datos de entrada.
Diciembre 2024	Sequence length too short for LSTM.	Longitud de las secuencias insuficiente para capturar tendencias.	Se aumentó la ventana temporal (sequence length).
Enero 2024	Error: Out of memory.	Red demasiado grande para la RAM disponible.	Se redujo el tamaño del batch y se entrenó en GPU.
Enero 2024	Unexpected NaN values in loss function.	Alguna transformación generaba valores no numéricos.	Se verificó que X_train y Y_train no tengan NaN (isnan()).
Enero 2024	Error using trainNetwork: Inputs must have consistent time steps.	Diferencias en la longitud de secuencias dentro del mismo batch.	Se utilizó padsequence() para igualar longitudes.
Enero 2024	Error: SequenceInputLayer requires 3D input.	Datos no tenían la forma esperada (num_features, num_timesteps, num_samples).	Se reorganizaron los datos con reshape() para que tengan la forma correcta.
Enero 2024	Error using trainNetwork: The size of X and Y must match.	Desfase en la sincronización de los datos de entrada y salida.	Se ajustó la alineación de X_train y Y_train considerando los lags.