



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA



**PREDICCIÓN DEL CONSUMO DE ENERGÍA MENSUAL EN LA
CIUDAD DE PUNO UTILIZANDO EL ALGORITMO RANDOM
FOREST, ELECTRO PUNO S.A.A. PERIODO 2005 – 2024**

TESIS

PRESENTADA POR:

Bach. EDSON DAVID MAMANI ALAVE

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2026



EDSON DAVID MAMANI ALAVE

PREDICCIÓN DEL CONSUMO DE ENERGÍA MENSUAL EN LA CIUDAD DE PUNO UTILIZANDO EL ALGORITMO RANDOM FO...

Universidad Nacional del Altiplano

Detalles del documento

Identificador de la entrega

trn:oid::8254:576135250

Fecha de entrega

8 abr 2026, 10:15 a.m. GMT-5

Fecha de descarga

8 abr 2026, 10:20 a.m. GMT-5

Nombre del archivo

TESIS_repo.docx

Tamaño del archivo

1.1 MB

69 páginas

11.486 palabras

67.180 caracteres



13% Similitud general

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

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Texto mencionado
- ▶ Coincidencias menores (menos de 12 palabras)

Fuentes principales

- 10% Fuentes de Internet
- 4% Publicaciones
- 8% Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

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

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



José P. Tito Lipa
Ing. Estadístico e Informático D.Sc.
CIP. 159645





Dr. BERNABE CANQUI FLORES
Director Unidad de Investigación
FINESI - UNA - PUNO





UNIVERSIDAD NACIONAL DEL ALTIPLANO - PUNO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

**PREDICCIÓN DEL CONSUMO DE ENERGÍA MENSUAL EN LA CIUDAD DE
PUNO UTILIZANDO EL ALGORITMO RANDOM FOREST, ELECTRO PUNO
S.A.A. PERIODO 2005 – 2024**

**TESIS PRESENTADA POR:
EDSON DAVID MAMANI ALAVE**

**PARA OPTAR EL TITULO PROFESIONAL DE:
INGENIERO ESTADÍSTICO E INFORMÁTICO**



APROBADA POR EL JURADO REVISOR CONFORMADO POR:

PRESIDENTE:


Dr. GODOFREDO QUISPE MAMANI

PRIMER MIEMBRO:


M. S.c. ROBERTO ELVIS ROQUE CLAROS

SEGUNDO MIEMBRO:


M. S.c. ALCIDES RAMOS CALCINA

DIRECTOR / ASESOR:


Dr. JOSE PANFILO TITO LIPA

ÁREA: Estadística Aplicada.

TEMA: Modelos Predictivos Uni y Multivariantes.

FECHA DE SUSTENTACIÓN: 06 de enero de 2026



DEDICATORIA

Dedico este trabajo a mis padres, David y Aurelia, por su amor incondicional, su constante sacrificio y por haberme brindado el regalo más valioso: la educación. A mis hermanos, por su apoyo sincero y su compañía a lo largo de cada etapa de mi vida.

A mis docentes, por sus enseñanzas, orientación académica, exigencia formativa y por compartir su conocimiento y experiencia, los cuales fueron fundamentales en mi proceso de aprendizaje y crecimiento profesional.

A mis amigos, por su apoyo incondicional, motivación constante y compañerismo, quienes hicieron más llevadero este camino y estuvieron presentes tanto en los momentos de esfuerzo como de logro.

Finalmente, a los profesionales que he tenido la oportunidad de conocer a lo largo de mi formación académica, por sus consejos, ejemplo y valioso aporte, los cuales contribuyeron significativamente a mi desarrollo académico, profesional y humano.

Edson David Mamani Alave



AGRADECIMIENTOS

Expreso mi sincero agradecimiento a mi Alma Mater, la Universidad Nacional del Altiplano de Puno, así como a los docentes de la Escuela Profesional de Ingeniería Expreso mi sincero agradecimiento a mi Alma Mater, la Universidad Nacional del Altiplano de Puno, y a los docentes de la Escuela Profesional de Ingeniería Estadística e Informática, por los conocimientos y aportes brindados durante mi formación profesional. De manera especial, agradezco a mi asesor de investigación, Dr. José Panfilo Tito Lipa, por su orientación y compromiso, fundamentales para el desarrollo y culminación del presente trabajo, así como a los miembros del jurado evaluador por sus valiosas observaciones y recomendaciones.

Agradezco a mi familia por su apoyo incondicional y motivación constante. Finalmente, extendiendo mi agradecimiento a todas las personas que, directa o indirectamente, contribuyeron a la realización y culminación de este estudio.

Edson David Mamani Alave



ÍNDICE GENERAL

	Pág
DEDICATORIA	
AGRADECIMIENTOS	
ÍNDICE GENERAL	
ÍNDICE DE TABLAS	
ÍNDICE DE FIGURAS	
ÍNDICE DE ANEXOS	
ACRÓNIMOS	
RESUMEN	14
ABSTRACT.....	15
CAPÍTULO I	
INTRODUCCIÓN	
1.1 PLANTEAMIENTO DEL PROBLEMA.....	17
1.1.1 Descripción del problema	17
1.2 FORMULACIÓN DE PROBLEMA	18
1.2.1 Problema General.....	18
1.2.2 Problemas Específicos	18
1.3 HIPÓTESIS DE LA INVESTIGACIÓN	19
1.3.1 Hipótesis General	19
1.3.2 Hipótesis Específicas	19
1.4 JUSTIFICACIÓN	19
1.5 OBJETIVOS DE LA INVESTIGACIÓN.....	20
1.5.1 Objetivo General	20
1.5.2 Objetivos Específicos	20



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1	ANTECEDENTES DE LA INVESTIGACIÓN	22
2.1.1	Nivel Internacional.....	22
2.1.2	Nivel Nacional	23
2.1.3	Nivel Local.....	25
2.2	MARCO TEÓRICO	27
2.2.1	Predicción.....	27
2.2.2	Series Temporales	27
2.2.3	Árboles de Decisión	28
2.2.4	Algoritmo Random Forest.....	29
2.2.5	Validación Cruzada	29
2.3	MARCO CONCEPTUAL	30
2.3.1	Energía Eléctrica	30
2.3.2	Machine Learning	30
2.3.3	Modelos Supervisados de Regresión	31

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1	UBICACIÓN GEOGRÁFICA	32
3.2	POBLACIÓN Y MUESTRA	32
3.2.1	Población.....	32
3.2.2	Muestra.....	32
3.3	RECOLECCIÓN DE DATOS	33
3.4	METODOLOGÍA DE LA INVESTIGACIÓN	33
3.4.1	Tipo de investigación	33



3.4.2	Nivel de Investigación.....	33
3.4.3	Enfoque de investigación	34
3.4.4	Metodología para la predicción de datos	34

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1	RESULTADOS.....	37
4.1.1	Limpieza y preparación de los datos	37
4.1.2	Análisis exploratorio de los datos	38
4.1.3	Modelado y distribución de los datos para el entrenamiento	42
4.1.4	Evaluación del modelo	44
4.1.5	Predicción del modelo	45
4.1.6	Predicción con el modelo ARIMA.....	47
4.1.6.1	Análisis de los componentes de la serie temporal	47
4.1.6.2	Estimación de parámetros	48
4.1.6.3	Identificación del modelo ARIMA.....	49
4.1.6.4	Evaluación del modelo ARIMA.....	51
4.1.6.5	Pronóstico con el modelo ARIMA	51
4.2	DISCUSIÓN	53
V.	CONCLUSIONES.....	56
VI.	RECOMENDACIONES	58
VII.	REFERENCIAS BIBLIOGRÁFICAS.....	60
ANEXOS.....		63



ÍNDICE DE TABLAS

	Pág
Tabla 1 Evolución Mensual de la Máxima Demanda (MW) del 2005 al 2013	40
Tabla 2 Evolución Mensual de la Máxima Demanda (MW) del 2014 al 2022	40
Tabla 3 Resultado del Cross-Validation con 10 fold	43
Tabla 4 Predicciones del modelo Random Forest para 2025	46
Tabla 5 Comparación con modelos alternativos (AIC y BIC)	50
Tabla 6 Métricas de validación de ARIMA	51
Tabla 7 Predicción del modelo ARIMA	52
Tabla 8 Comparación de métricas entre Random Forest y ARIMA	53



ÍNDICE DE FIGURAS

	Pág
Figura 1 Evolución de la Demanda Máxima (MW) 2005-2025	41
Figura 2 Validación del Modelo Random Forest con los datos de 2022 y 2025	44
Figura 3 Predicción de Random Forest 2005 - 2025	47
Figura 4 Predicción de ARIMA	50
Figura 5 Pronóstico del Modelo ARIMA	52



ÍNDICE DE ANEXOS

	Pág
Anexo 1. Gráfico de residuales de ARIMA	63
Anexo 2. Código de implementación de algoritmos Random Forest	63
Anexo 3. Declaración de autenticidad.....	68
Anexo 4. Autorización para el depósito de tesis en el Repositorio Institucional.....	69



ACRÓNIMOS

ARIMA:	Modelo Autorregresivo Integrado de Media Móvil
MAE:	Error Absoluto Medio
MAPE:	Error Porcentual Absoluto Medio
RMSE:	Raíz del Error Cuadrático Medio
MW:	Megavatio
S.A.A.:	Sociedad Anónima Abierta
R:	Lenguaje de programación R
CV:	Validación Cruzada



RESUMEN

El objetivo de esta investigación fue estimar y evaluar el consumo eléctrico mensual en la ciudad de Puno mediante el modelo de aprendizaje automático Random Forest, con el fin de determinar su capacidad predictiva y contribuir a una mejor planificación energética regional; el estudio fue de enfoque cuantitativo, aplicado y predictivo, utilizando registros mensuales de demanda del período 2005–2025 proporcionados por Electro Puno S.A.A. y procesados en el lenguaje de programación R. El análisis evidenció una tendencia creciente y patrones estacionales asociados al crecimiento poblacional, la expansión urbana y el incremento de las actividades económicas, los cuales fueron modelados eficazmente por el algoritmo Random Forest. Los resultados mostraron que este modelo superó al ARIMA, presenta menores errores de predicción $RMSE = 8.448$ y $MAE = 6.375$ frente a $RMSE = 12.837$ y $MAE = 11.397$, lo que confirma su mayor precisión y capacidad para capturar relaciones no lineales y comportamientos estacionales complejos. Las proyecciones para el año 2025 indican una tendencia creciente de la demanda eléctrica, con valores estimados entre 73 y 80 MW, información útil para la planificación de infraestructura y la gestión eficiente de los recursos energéticos. Se concluyó que el modelo Random Forest es una herramienta robusta y confiable para la predicción del consumo eléctrico en la ciudad de Puno, confirmando la hipótesis general de investigación; asimismo, se concluyó que la actualización periódica del modelo, la incorporación de variables económicas y demográficas, y la evaluación de otros algoritmos de aprendizaje automático fortalecerán la planificación energética regional y promoverán un uso más eficiente y sostenible de la energía.

Palabras clave: Consumo de Energía, Electro Puno S.A.A., Planificación Energética, Predicción, Random Forest.



ABSTRACT

The objective of this research was to estimate and evaluate monthly electricity consumption in the city of Puno using the Random Forest machine learning model, in order to determine its predictive capacity and contribute to improved regional energy planning. The study employed a quantitative, applied, and predictive approach, utilizing monthly demand records from 2005 to 2025 provided by Electro Puno S.A.A. and processed using the R programming language. The analysis revealed an increasing trend and seasonal patterns associated with population growth, urban expansion, and increased economic activity, all of which were effectively modeled by the Random Forest algorithm. The results showed that this model outperformed ARIMA, exhibiting lower prediction errors (RMSE = 8.448 and MAE = 6.375) compared to ARIMA's RMSE = 12.837 and MAE = 11.397, confirming its greater accuracy and ability to capture nonlinear relationships and complex seasonal behaviors. Projections for 2025 indicate a growing trend in electricity demand, with estimated values between 73 and 80 MW. This information is useful for infrastructure planning and the efficient management of energy resources. It was concluded that the Random Forest model is a robust and reliable tool for predicting electricity consumption in the city of Puno, confirming the general research hypothesis. Furthermore, it was concluded that periodic model updates, the incorporation of economic and demographic variables, and the evaluation of other machine learning algorithms will strengthen regional energy planning and promote more efficient and sustainable energy use.

Keywords: Energy Consumption, Electro Puno S.A.A., Energy Planning, Prediction, Random Forest.



CAPÍTULO I

INTRODUCCIÓN

El consumo de electricidad es uno de los pilares fundamentales del desarrollo económico, industrial y social de cualquier región, ya que está directamente vinculado a la productividad, el crecimiento urbano, la modernización tecnológica y la mejora de la calidad de vida. En el caso de la ciudad de Puno, el crecimiento poblacional, el aumento de la actividad comercial y la expansión de los servicios públicos han generado una demanda energética cada vez más alta y variable. Este aumento sostenido plantea importantes desafíos para las empresas distribuidoras y los organismos de planificación energética, que deben anticipar los patrones de consumo para garantizar un suministro continuo, eficiente y sostenible. La falta de pronósticos adecuados del consumo puede generar desequilibrios entre la oferta y la demanda, generando pérdidas económicas, interrupciones del servicio e impactos negativos en la productividad regional. Por lo tanto, los modelos predictivos precisos y fiables se han convertido en una herramienta esencial para optimizar la gestión eléctrica y promover el desarrollo energético planificado en Puno y sus alrededores.

El objetivo principal de esta investigación es predecir el consumo eléctrico mensual de la ciudad de Puno mediante el algoritmo de aprendizaje automático Random Forest, aplicando técnicas modernas de aprendizaje automático que superan las limitaciones de los modelos estadísticos tradicionales. Para ello, se analizaron datos históricos de demanda máxima mensual de 2005-2025, proporcionados por Electro Puno S.A.A., para proyectar la demanda esperada para 2025. El modelo se desarrolló y ejecutó utilizando el lenguaje de programación R, debido a su eficiencia en el manejo y procesamiento de grandes volúmenes de datos, así como a sus amplias capacidades de



análisis estadístico, visualización y modelado predictivo. Se utilizaron bibliotecas especializadas como randomForest y caret, que permitieron construir, refinar y validar el modelo mediante técnicas de validación cruzada, obteniendo así resultados robustos y reproducibles. El estudio busca demostrar que el modelo Random Forest es capaz de capturar relaciones no lineales y tendencias estacionales presentes en los datos de consumo energético, proporcionando así una estimación más precisa y fiable que modelos convencionales como ARIMA. De esta forma, la investigación no solo contribuye al ámbito técnico y científico, sino que también ofrece una herramienta útil para la toma de decisiones estratégicas en la planificación energética, beneficiando a la población, las empresas y las instituciones locales.

1.1 PLANTEAMIENTO DEL PROBLEMA

1.1.1 Descripción del problema

El consumo de energía eléctrica es un factor determinante en el desarrollo económico y social de las ciudades, al incidir directamente en las actividades industriales, comerciales y residenciales. En la ciudad de Puno, la demanda energética ha crecido de manera sostenida como resultado del incremento poblacional, la expansión urbana y la incorporación de nuevas tecnologías. No obstante, la planificación del sistema de distribución eléctrica enfrenta importantes desafíos debido a la variabilidad del consumo, la presencia de estacionalidad y posibles cambios estructurales en la serie. Electro Puno S.A.A., empresa responsable del suministro eléctrico en la región, dispone de registros históricos de consumo; sin embargo, se requiere de herramientas analíticas avanzadas para el aprovechamiento de esta información y la toma de decisiones estratégicas.



Tradicionalmente, el pronóstico de la demanda eléctrica se ha abordado mediante modelos estadísticos de series temporales, destacando el modelo ARIMA, el cual requiere el cumplimiento de supuestos previos como la estacionariedad, la linealidad y la independencia de los errores. Sin embargo, el consumo eléctrico suele presentar patrones no lineales, estacionalidad compleja y variaciones estructurales que pueden afectar el desempeño de estos modelos. En este contexto, los enfoques basados en aprendizaje automático, particularmente el algoritmo Random Forest, surgen como una alternativa capaz de capturar relaciones no lineales y patrones complejos sin depender estrictamente de dichos supuestos. Por ello, la presente investigación propone comparar el desempeño predictivo de los modelos ARIMA y Random Forest en la estimación del consumo eléctrico en la ciudad de Puno, con el fin de determinar cuál proporciona pronósticos más precisos y confiables para mejorar la planificación energética y la gestión eficiente de los recursos.

1.2 FORMULACIÓN DE PROBLEMA

1.2.1 Problema General

¿Cómo estimar con precisión el consumo mensual de energía eléctrica en la ciudad de Puno, utilizando técnicas de machine learning como el algoritmo Random Forest, de manera que se pueda mejorar la planificación energética y garantizar un uso eficiente de los recursos disponibles?

1.2.2 Problemas Específicos

- ¿Qué patrones y tendencias se identifican en los datos históricos de consumo de energía eléctrica de Puno en el periodo 2005 – 2025?

- ¿Qué nivel de precisión y confiabilidad presenta el modelo Random Forest al predecir el consumo energético, medido mediante métricas de evaluación como RMSE y MAE?
- ¿En qué medida el modelo Random Forest supera al modelo ARIMA en la predicción del consumo mensual de energía en la ciudad de Puno?

1.3 HIPÓTESIS DE LA INVESTIGACIÓN

1.3.1 Hipótesis General

El modelo Random Forest sí permitió estimar con precisión el consumo de energía en la ciudad de Puno para el periodo 2005-2025.

1.3.2 Hipótesis Específicas

- El comportamiento del consumo de energía en la ciudad de Puno presentó una tendencia creciente debido al aumento de la población y el desarrollo económico, sin variaciones cíclicas y estacionales significativas.
- Las métricas de evaluación RMSE y MAE demostraron un rendimiento predictivo eficiente y confiable en el modelo de Random Forest.
- El modelo Random Forest presentó un mejor rendimiento en la predicción del consumo energético en comparación con el modelo ARIMA.

1.4 JUSTIFICACIÓN

El consumo de energía es un tema crucial para la población de Puno. Por lo tanto, resulta fundamental considerar las mediciones futuras del consumo eléctrico. El pronóstico debía ser lo más realista posible, ya que variaciones significativas con respecto a los datos reales podrían provocar pérdidas económicas y fallas en el servicio brindado por Electro Puno S.A.A. En los



últimos años, el servicio eléctrico ha experimentado un incremento en los costos, asociado principalmente al crecimiento poblacional e industrial.

Para analizar y proyectar el consumo eléctrico, se aplicó el algoritmo Random Forest, una metodología avanzada de aprendizaje automático. Esta técnica permitió describir, analizar y pronosticar el historial de consumo energético de Puno con mayor precisión y flexibilidad. Asimismo, en comparación con modelos estadísticos tradicionales como ARIMA, el modelo Random Forest presenta una menor complejidad en sus etapas de implementación, ya que no requiere supuestos estrictos de estacionariedad ni procesos extensos de identificación y ajuste de parámetros, lo que facilita su ejecución y aplicación práctica. Para mejorar el servicio a la población de Puno, se desarrolló un modelo de pronóstico ajustado al consumo eléctrico mensual, lo que permitió un acceso más eficiente a la electricidad, beneficiando a los sectores industrial, comercial y residencial, e impulsando su desarrollo.

1.5 OBJETIVOS DE LA INVESTIGACIÓN

1.5.1 Objetivo General

Estimar y evaluar el consumo de energía en la ciudad de Puno utilizando el modelo Random Forest, con el fin de identificar su capacidad predictiva para los próximos años.

1.5.2 Objetivos Específicos

- Analizar los datos históricos del consumo de energía en Puno desde el año 2005 hasta 2025, con el fin de identificar tendencias y patrones relevantes.



- Evaluar la capacidad predictiva del modelo Random Forest mediante la validación cruzada y métricas de error como RMSE y MAE, aplicadas a los datos históricos.
- Comparar el desempeño del modelo Random Forest con el modelo ARIMA.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES DE LA INVESTIGACIÓN

2.1.1 Nivel Internacional

Torres Sánchez (2023) desarrolló un estudio orientado al pronóstico de la demanda de energía eléctrica horaria en el departamento de Antioquia, Colombia, considerando un horizonte de corto plazo de una semana con intervalos de una hora. La investigación incluyó procesos de extracción y limpieza de datos, análisis exploratorio de series temporales y la implementación de modelos predictivos basados en métodos estadísticos clásicos, algoritmos de machine learning autorregresivos y técnicas de deep learning enfocadas en series de tiempo. Los resultados evidenciaron que los modelos de aprendizaje automático y profundo presentaron un mejor desempeño predictivo frente a los métodos tradicionales, al capturar de manera más eficiente patrones no lineales y variaciones temporales complejas, contribuyendo a la reducción de sobrecostos operativos y al fortalecimiento de la confiabilidad del suministro eléctrico, lo cual sustenta la aplicabilidad de estas técnicas en la predicción de la generación de energía solar fotovoltaica.

Rojas Victoriano (2022) desarrolló una memoria para optar al título de Ingeniero Civil Eléctrico titulada Diseño e implementación de modelos de intervalos neuronales para sistema de gestión de agua-energía de cultivos, en la cual abordó el problema del impacto de las variables climatológicas en sistemas de predicción orientados a la gestión integrada de agua y energía en una



comunidad rural de Chile. La investigación se centró en la cuantificación de la incertidumbre asociada a la predicción de variables climáticas como radiación solar, temperatura del aire, precipitaciones y humedad relativa, con el fin de utilizarlas como entradas de un sistema de control agua-energía (EWMS) en una microred comunitaria. Para ello, el estudio contempló la recopilación y validación de datos provenientes de una estación meteorológica local y de fuentes externas, el diseño e implementación de modelos de predicción basados en redes neuronales (perceptrón multicapa y LSTM), así como el desarrollo de modelos de intervalos de predicción mediante enfoques Joint Supervision y Quality Driven. Los resultados evidenciaron un desempeño aceptable en la mayoría de las variables analizadas, destacando la importancia de considerar la incertidumbre en los modelos predictivos para sistemas energéticos, aunque se identificaron limitaciones en la predicción de precipitaciones diarias, lo que sugiere la necesidad de continuar investigando metodologías más robustas para este tipo de variables.

2.1.2 Nivel Nacional

Vargas Zavala (2025) desarrolló una tesis titulada “Desarrollar un modelo de proyección de demanda eléctrica diaria del Sistema Eléctrico Interconectado Nacional (SEIN) utilizando Redes Neuronales Recurrentes y Transformada Wavelet”, en la cual abordó la problemática de la proyección precisa de la demanda eléctrica en el mercado eléctrico peruano, considerando su impacto en la optimización del consumo de los clientes libres y en la planificación del despacho de los generadores. La investigación propuso un modelo híbrido que combinó Redes Neuronales Recurrentes del tipo Long Short Term Memory (LSTM) con la Transformada Wavelet Discreta (DWT), técnica que permitió



descomponer la señal de demanda eléctrica en múltiples sub-señales asociadas a diferentes componentes de frecuencia, reduciendo la volatilidad y facilitando el análisis de patrones temporales. Cada sub-señal fue procesada mediante modelos LSTM y posteriormente reconstruida utilizando la Transformada Wavelet Estacionaria Inversa, preservando la longitud original de la serie temporal. El estudio empleó datos de demanda eléctrica con una resolución de quince minutos correspondientes al período enero de 2021 a noviembre de 2023, obtenidos del Comité de Operación Económica del Sistema (COES). Los resultados evidenciaron que el modelo híbrido propuesto superó en precisión tanto al modelo LSTM tradicional como al modelo de predicción diaria utilizado por el COES, alcanzando mejores indicadores de desempeño estadístico, lo que demostró la efectividad de la integración de técnicas wavelet y redes neuronales recurrentes para la proyección de la demanda eléctrica diaria del SEIN.

Pinedo Amias & Lozano Gomez (2022) en su proyecto de tesis titulado “Modelo De Red Neuronal Para La Predicción Del Consumo De Energía Eléctrica En Iquitos” Desarrollaron un modelo de red neuronal artificial para la predicción del consumo de energía eléctrica en la ciudad de Iquitos, Su investigación tuvo como objetivo principal evaluar la implementación de un modelo de red neuronal artificial para mejorar la capacidad de pronóstico del consumo eléctrico en la ciudad de Iquitos. La metodología que utilizó en la investigación se conformó por la recopilación de datos históricos sobre el consumo de electricidad en Iquitos, la implementación de un modelo de red neuronal artificial y la evaluación de la capacidad de pronóstico del modelo utilizando métricas estadísticas como el coeficiente de correlación (R) y el error cuadrático medio (RMSE). La hipótesis de su investigación sugiere que la implementación de un modelo de red neuronal



artificial permitirá mejorar la capacidad de pronóstico del consumo eléctrico en Iquitos, lo que contribuirá a una mejor gestión y eficiente de la electricidad en la Iquitos.

García Fernández (2021) en su tesis titulada “Modelamiento Del Pronóstico De La Demanda Eléctrica Diaria Del Sistema Eléctrico Interconectado Nacional Utilizando Técnicas De Machine Learning” El estudio se centró en la comparación de dos metodologías: redes neuronales ARTMAP Fuzzy y redes ANFIS Neuro-Fuzzy, evaluando su desempeño con preprocesamiento de datos para mejorar la precisión en los resultados. La investigación, realizada con Matlab R2016a, propuso optimizar el Procedimiento Técnico 03 del COES para reducir el error porcentual absoluto medio y mejorar la gestión de recursos energéticos. Los resultados destacaron la viabilidad de estas técnicas para optimizar la programación diaria del despacho hidrotérmico y garantizar una planificación energética más eficiente.

2.1.3 Nivel Local

Añasco Chata (2023) realizó un modelo de series de tiempo utilizando la metodología Box-Jenkins para pronosticar el consumo de energía eléctrica en la región de Puno. El estudio se basó en datos históricos mensuales de consumo desde 2014 hasta 2020, proporcionados por ELECTRO PUNO S.A.A. El modelo multiplicativo estacional ARIMA $(0,1,3) \times (2,0,0)_{12}$ se identificó como el mejor ajuste para la serie, con una alta precisión de pronóstico. Sin embargo, se concluyó que eventos mundiales, como la pandemia del año 2020, influyen negativamente a la serie de consumo de energía eléctrica, afectando la identificación y precisión del modelo ARIMA, ya que estos modelos dependen de los datos más recientes



de la serie, lo que desfavorece la exactitud de los pronósticos. Basándonos en estos antecedentes, el presente proyecto propone la utilización del algoritmo de Random Forest para para investigaciones futuras sobre predicciones de consumo de energía eléctrica en la ciudad de Puno.

Suxso Marca (2024) desarrolló la tesis titulada “Predicción de la generación de energía eléctrica de un sistema fotovoltaico mediante Machine Learning en la región Puno”, con el objetivo de mejorar la planificación y gestión de los recursos energéticos a partir de la estimación de la producción de energía fotovoltaica en un contexto de alta variabilidad climática. La investigación se centró en el diseño y entrenamiento de un modelo basado en redes neuronales de memoria a largo plazo (LSTM), empleando exclusivamente variables eléctricas como corriente, voltaje y potencia, sin incorporar información climática adicional. Metodológicamente, se implementó un sistema de monitoreo para la recolección de datos del sistema fotovoltaico, los cuales fueron preprocesados y divididos en conjuntos de entrenamiento y prueba en proporciones de 80% y 20%, respectivamente. El desempeño del modelo fue evaluado mediante métricas estadísticas como el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE). Los resultados evidenciaron un buen nivel de precisión, alcanzando un R^2 de 0.90, lo que demostró la capacidad del modelo para explicar la mayor parte de la variabilidad en la generación de energía eléctrica a partir de datos puramente eléctricos, concluyéndose que los enfoques de Machine Learning constituyen una alternativa viable para la predicción de la generación fotovoltaica, aunque se recomendó la inclusión de variables climáticas en estudios futuros para mejorar la precisión del modelo.



2.2 MARCO TEÓRICO

2.2.1 Predicción

La previsión constituye un proceso fundamental en la toma de decisiones estratégicas y operativas de las organizaciones, pues su propósito principal es anticipar, con la mayor precisión posible, el comportamiento futuro de una variable a partir de la información disponible en el presente. Para ello, se consideran no solo los datos históricos, sino también el conocimiento de posibles eventos futuros que puedan incidir en los resultados esperados. Según (Rob J. Hyndman y George Athanasopoulos (2018), prever implica desarrollar métodos que permitan minimizar la incertidumbre y generar proyecciones confiables, lo que convierte al pronóstico en una herramienta esencial para actividades como la planificación económica, la gestión de recursos, el control de inventarios y la optimización de procesos.

No obstante, el pronóstico no se limita únicamente a la aplicación mecánica de modelos matemáticos. Requiere de un sistema integral que involucre la identificación clara del problema de predicción, la selección de técnicas adecuadas para cada contexto y la evaluación continua de su desempeño. (Rob J. Hyndman y George Athanasopoulos, 2018) destacan que un sistema de pronóstico efectivo demanda la combinación de conocimientos técnicos especializados con un respaldo organizacional sólido, de modo que los resultados obtenidos se integren en la toma de decisiones.

2.2.2 Series Temporales

Una serie temporal puede entenderse como un conjunto de observaciones recolectadas de manera secuencial en el tiempo, lo que permite analizar cómo



evoluciona una variable en un determinado periodo. Ejemplos comunes de series temporales incluyen los precios diarios de acciones de empresas como IBM, las precipitaciones registradas mensualmente en una localidad, las ventas trimestrales de corporaciones como Amazon o los beneficios anuales reportados por compañías como Google. En todos estos casos, lo fundamental es que las observaciones mantienen un orden cronológico, lo que la diferencia de otros tipos de datos Rob J. Hyndman y George Athanasopoulos (2018)

En este sentido, Box, Jenkins, Reinsel y Ljung (2016) explican que la previsión en series temporales consiste en utilizar las observaciones disponibles hasta un tiempo t para estimar valores futuros en un horizonte temporal $t + l$. Este enfoque ha sido aplicado tradicionalmente en áreas como la planificación económica y empresarial, la producción, la gestión de inventarios y el control de procesos industriales. Su aporte radica en que permite estructurar de manera sistemática el uso de datos históricos para enfrentar la incertidumbre y facilitar la optimización de los recursos.

2.2.3 Árboles de Decisión

Los árboles de decisión constituyen una de las metodologías más empleadas en el campo del aprendizaje automático para la clasificación y predicción de datos. De acuerdo con J. Ross Quinlan (1993), un árbol de decisión es una estructura jerárquica que puede estar formada por dos tipos de nodos:

- Nodos hoja, que representan una clase o resultado final.
- Nodos de decisión, que contienen una prueba sobre un atributo específico de los datos, generando ramas según los posibles valores que dicho atributo puede asumir.



2.2.4 Algoritmo Random Forest

El algoritmo Random Forest es una técnica de *ensemble learning* desarrollada por Breiman (2001), que se basa en la construcción de múltiples árboles de decisión y la combinación de sus resultados para obtener un modelo predictivo más robusto y preciso. Su fundamento principal proviene de la extensión del método de *bagging* (bootstrap aggregating), con la diferencia clave de que, en cada división de los árboles, se selecciona aleatoriamente un subconjunto de predictores en lugar de considerar todas las variables disponibles. Este procedimiento reduce la correlación entre los árboles, generando una mayor diversidad en el conjunto y mejorando la capacidad de generalización del modelo.

De acuerdo con Kuhn & Johnson (2013), Random Forest puede entenderse como un conjunto de árboles de decisión “decorrelacionados”, cuyo objetivo es disminuir la varianza de las predicciones y aumentar la estabilidad del modelo frente a datos ruidosos o altamente variables. Cada árbol contribuye con una predicción individual y el resultado final se obtiene mediante votación (en problemas de clasificación) o mediante el promedio (en problemas de regresión).

2.2.5 Validación Cruzada

La validación cruzada constituye uno de los métodos más empleados para la estimación y evaluación de modelos predictivos, así como para la selección de variables. De acuerdo con (Jung & Hu, 2015), el procedimiento más utilizado es el *k-fold cross validation*, que consiste en dividir el conjunto de datos en *k* pliegues o subconjuntos de igual tamaño, entrenando el modelo en *k-1* pliegues y validándolo en el pliegue restante. Este proceso se repite *k* veces y finalmente se



obtiene una estimación promedio del error de predicción, lo que proporciona una medida más objetiva y estable del rendimiento del modelo.

2.3 MARCO CONCEPTUAL

2.3.1 Energía Eléctrica

La energía eléctrica puede definirse como la forma de energía asociada al movimiento de electrones a través de un conductor, originado por la existencia de una diferencia de potencial eléctrico entre dos puntos. Este desplazamiento de cargas genera la corriente eléctrica, la cual puede transformarse en otras formas útiles de energía, tales como luz, calor o energía mecánica, que son aprovechadas en múltiples aplicaciones de la vida cotidiana, la industria y los sistemas de comunicación.

La energía eléctrica se considera una energía secundaria, dado que no se encuentra de manera directa en la naturaleza, sino que es producida a partir de diversas fuentes de energía primaria. De acuerdo con Repsol (2025), estas incluyen recursos convencionales, como los combustibles fósiles (carbón, gas natural y petróleo) y la energía nuclear, así como fuentes renovables, como la energía eólica, solar, hidráulica y geotérmica. Su carácter versátil, limpio en el punto de consumo y fácilmente transportable mediante redes de transmisión y distribución la convierten en uno de los pilares fundamentales del desarrollo económico y social contemporáneo.

2.3.2 Machine Learning

El Machine Learning puede entenderse como un proceso de búsqueda dentro de un extenso espacio de hipótesis, con el fin de identificar aquella que



mejor se ajuste a los datos observados y al conocimiento previo disponible. Según Mitchell (1997), cada hipótesis representa un modelo posible que describe la relación entre las variables de entrada y salida; por ejemplo, en el caso de un sistema que aprende a jugar damas, el espacio de hipótesis estaría conformado por todas las funciones de evaluación que resultan de diferentes combinaciones de parámetros o pesos asignados. La tarea fundamental del aprendiz consiste en explorar este vasto conjunto de posibilidades para localizar la hipótesis que muestre la mayor consistencia con los datos, equilibrando precisión predictiva y coherencia con las restricciones iniciales.

2.3.3 Modelos Supervisados de Regresión

El aprendizaje supervisado es una de las principales categorías del aprendizaje estadístico y se caracteriza porque cada observación del conjunto de datos contiene tanto un vector de variables predictoras como una variable de respuesta asociada, lo que permite entrenar modelos capaces de establecer relaciones entre ambas. De acuerdo con James et al., (2021) ,el propósito central de este enfoque es predecir con precisión la respuesta para nuevas observaciones o comprender mejor los vínculos entre predictores y respuesta. Entre los métodos supervisados tradicionales destacan la regresión lineal y la regresión logística, mientras que dentro de las técnicas modernas se incluyen los modelos aditivos generalizados, el boosting y las máquinas de vectores de soporte. En todos los casos, el modelo se entrena con datos etiquetados, lo que le otorga gran utilidad en tareas de predicción e inferencia en múltiples campos de aplicación.



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 UBICACIÓN GEOGRÁFICA

La investigación se realizó en la ciudad de Puno en el ámbito de servicio de la empresa Electro Puno S.A.A., en el Distrito Puno, Provincia de Puno, Región Puno, País Perú.

3.2 POBLACIÓN Y MUESTRA

3.2.1 Población

La población estuvo constituida por los registros históricos mensuales del consumo de energía eléctrica en Mega watts de la ciudad de Puno, registrados por la empresa Electro Puno S.A.A., en los periodos comprendidos entre los años 2005-2025.

3.2.2 Muestra

La muestra estuvo compuesta por todos los registros mensuales del consumo de energía eléctrica en la ciudad de Puno durante el periodo 2005-2025. Se utiliza la totalidad de los datos disponibles, ya que representan de forma completa la evolución del consumo energético, permitiendo un análisis preciso y una adecuada base para el entrenamiento y validación del modelo predictivo.



3.3 RECOLECCIÓN DE DATOS

Se recopilaron datos históricos del consumo de energía previstas en las memorias anuales publicadas hasta la fecha por la empresa Electro Puno S.A.A. en la ciudad de Puno desde 2005-2025.

3.4 METODOLOGÍA DE LA INVESTIGACIÓN

3.4.1 Tipo de investigación

La presente investigación es aplicada, dado que busca ofrecer una solución práctica a un problema real vinculado con la planificación energética en la ciudad de Puno, mediante la utilización de técnicas de predicción basadas en aprendizaje automático.

De acuerdo con Hernández Sampieri et al., (2014), este tipo de investigación conserva el rigor científico propio del método es sistemática, empírica y crítica, pero su finalidad es transformar la realidad o mejorar procesos existentes a través de la generación de resultados útiles y aplicables. Es decir, combina teoría y práctica para intervenir en situaciones reales, proporcionando alternativas o mejoras basadas en evidencia.

3.4.2 Nivel de Investigación

Predictivo: Este nivel se orienta a la construcción de modelos matemáticos que permiten estimar la probabilidad de ocurrencia de determinados fenómenos a partir de datos observados previamente. Según José Supo (2025), el nivel de investigación predictivo busca anticipar hechos o comportamientos futuros mediante el uso de modelos estadísticos o algoritmos, partiendo del conocimiento ya existente sobre las causas que originan dichos fenómenos. En este estudio, se

aplica este nivel porque el objetivo principal es anticipar el comportamiento del consumo de energía eléctrica en la ciudad de Puno, utilizando datos históricos proporcionados por Electro Puno S.A.A. y aplicando técnicas de aprendizaje automático, específicamente el algoritmo Random Forest, para identificar patrones y realizar predicciones con base en ellos.

3.4.3 Enfoque de investigación

El presente estudio se enmarca dentro del enfoque cuantitativo, ya que se sustenta en la obtención y el análisis numérico de datos históricos sobre el consumo de energía eléctrica en la ciudad de Puno. A través de este enfoque se busca identificar patrones, establecer relaciones y realizar predicciones mediante procedimientos estadísticos y el uso de algoritmos de aprendizaje automático.

De acuerdo con Hernández Sampieri et al. (2014), el enfoque cuantitativo se caracteriza por apoyarse en la medición y el análisis estadístico de los datos con el propósito de probar hipótesis, formular inferencias objetivas y confirmar teorías. Este enfoque sigue un proceso estructurado y secuencial que parte de la delimitación de una idea o problema, la definición de variables e hipótesis, y culmina con la interpretación de resultados a partir de los datos recolectados.

3.4.4 Metodología para la predicción de datos

Para el desarrollo de la predicción de datos se utilizó el modelo Random Forest el cual combina múltiples árboles de decisión, donde cada árbol se basa en un vector aleatorio de características seleccionado de forma independiente. La precisión del modelo depende de la fuerza individual de cada árbol y de la correlación entre ellos. La aleatoriedad en la selección de características durante las divisiones de nodos genera tasas de error competitivas con técnicas como

AdaBoost, pero con mayor robustez frente al ruido. Además, los métodos internos permiten estimar errores, importancia de variables y evaluar cómo influyen las características seleccionadas en el modelo, aplicándose tanto a problemas de clasificación como de regresión Breiman, (2001).

Validación: En la validación Cruzada el total de los datos se dividen en k subconjuntos, de manera que aplicamos el método hold-out k veces, utilizando cada vez un subconjunto distinto para validar el modelo entrenado con los otros $k-1$ subconjuntos. El error medio obtenido de los k análisis realizados nos proporciona el error cometido por el método, permitiendo así evaluar su validez (Jung & Hu, 2015).

Métricas de Evaluación: El modelo fue evaluado utilizan el Error Cuadrático Medio (RMSE) como también el Error Absoluto Medio (MAE).

El error cuadrático medio (RMSE) y el error absoluto medio (MAE) son dos métricas estándar utilizadas en la evaluación de modelos. Para una muestra de n observaciones y n predicciones del modelo correspondiente.

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

n : Número total de observaciones.

i : Índice de cada observación.

y_i : Valor real u observado de la variable en la observación i .

\hat{y}_i : Valor estimado o predicho por el modelo para la observación i .

$y_i - \hat{y}_i$: Error de predicción en la observación i .



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

n : Número total de observaciones.

i : Índice de cada observación.

y_i : Valor real u observado de la variable en la observación i .

\hat{y}_i : Valor estimado o predicho por el modelo para la observación i .

$|y_i - \hat{y}_i|$: Valor absoluto del error de predicción en la observación i .

Como su nombre lo indica, el RMSE es la raíz cuadrada del error cuadrático medio (MSE) (Hodson, 2022).



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 RESULTADOS

Se implementó el modelo de aprendizaje automático Random Forest para predecir la demanda máxima mensual de electricidad en la ciudad de Puno. Se utilizaron datos históricos de 2005 a 2025, proporcionados por Electro Puno S.A.A., para estimar las tendencias de demanda esperadas para 2025. El modelo se desarrolló y ejecutó en el lenguaje de programación R, gracias a su eficiente gestión de datos, su capacidad para aplicar algoritmos de aprendizaje automático y sus herramientas de análisis estadístico y visualización de resultados.

4.1.1 Limpieza y preparación de los datos

Para la aplicación del algoritmo Random Forest, inicialmente se realizó la limpieza y preparación de los datos históricos de la demanda máxima mensual de energía eléctrica correspondientes al periodo 2005–2025, proporcionados por Electro Puno S.A.A. Los registros fueron organizados en una base de datos estructurada que incluyó las variables año, mes y demanda máxima mensual expresada en megavatios (MW).

Se verificó la consistencia de los datos, comprobando la ausencia de valores faltantes y errores de registro. Posteriormente, la variable mes fue transformada en una variable categórica (factor), con el propósito de permitir que el modelo identifique adecuadamente los patrones estacionales presentes en la serie. Esta etapa garantizó que la información utilizada fuera coherente y adecuada para el proceso de modelado predictivo.

4.1.2 Análisis exploratorio de los datos

Una vez depurados los datos, se realizó un análisis exploratorio con el objetivo de caracterizar el comportamiento de la demanda máxima mensual de energía eléctrica en la ciudad de Puno durante el periodo 2005–2025. La inspección visual de la serie temporal (Figura 1) permite identificar una tendencia creciente de largo plazo, evidenciando un aumento sostenido del consumo eléctrico asociado al crecimiento poblacional, la expansión de la actividad económica y la mayor penetración de tecnologías eléctricas en los sectores residencial, comercial e institucional. Este comportamiento confirma que la demanda energética en la ciudad presenta una dinámica no estacionaria, lo cual constituye un aspecto relevante para la selección de modelos de pronóstico.

Adicionalmente, se observan fluctuaciones estacionales claramente definidas, con picos recurrentes de consumo entre los meses de junio y octubre, y mínimos sistemáticos durante los meses de enero y febrero. Este patrón sugiere la presencia de una estacionalidad anual asociada a factores climáticos, sociales y económicos propios de la región, tales como el aumento de actividades productivas en determinados meses, el calendario académico y variaciones en las necesidades de calefacción e iluminación. La persistencia de este comportamiento a lo largo de los años indica que la estacionalidad constituye un componente estructural de la serie y no un fenómeno aleatorio.

El análisis de las Tablas 1 y 2 permite cuantificar esta evolución. Entre el periodo 2005-2013 se observa un crecimiento progresivo de la demanda máxima mensual, pasando de valores cercanos a 33 MW en los primeros años a niveles superiores a 60 MW hacia 2013. Este incremento se mantiene en el periodo 2014–



2019, alcanzando máximos cercanos a 80 MW, lo que refleja una expansión sostenida del consumo eléctrico en la ciudad. La tendencia positiva de largo plazo confirma que la demanda no puede considerarse estacionaria en media, condición fundamental exigida por los modelos ARIMA tradicionales.

Un aspecto relevante del análisis corresponde al impacto del contexto excepcional generado por la pandemia de COVID-19 en los años 2020 y 2021. Durante este periodo se observa una alteración significativa del patrón histórico de consumo, evidenciada por variaciones atípicas en la demanda máxima mensual. En particular, en 2020 se registran valores inusuales en algunos meses, como la marcada reducción en noviembre (52.73 MW), que contrasta con los niveles superiores a 80 MW observados en meses cercanos. Este comportamiento puede atribuirse a la paralización parcial de actividades económicas, restricciones de movilidad, cierre de establecimientos comerciales e industriales y cambios en los hábitos de consumo derivados del confinamiento y el trabajo remoto. Si bien el consumo residencial pudo incrementarse, la contracción de la actividad productiva y comercial generó una reducción neta en la demanda agregada.

Posteriormente, durante 2021 y 2022, se evidencia un proceso de recuperación progresiva del consumo, aunque con una mayor variabilidad respecto a los años previos a la pandemia. Esta presencia de choques externos y cambios estructurales en la serie constituye una limitación importante para los modelos estadísticos clásicos como ARIMA, los cuales asumen estabilidad en la estructura temporal y relaciones lineales. La pandemia introduce rupturas en la tendencia y en los patrones de estacionalidad, afectando la capacidad de estos modelos para adaptarse a contextos no previstos en los datos históricos.

La Figura 1 muestra la evolución de la demanda máxima mensual de energía eléctrica entre los años 2005-2025, permitiendo visualizar claramente la tendencia y los ciclos estacionales de la serie.

Tabla 1

Evolución Mensual de la Máxima Demanda (MW) del 2005 al 2014

MES	AÑO									
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Enero	33.23	33.88	35.48	39.44	42.17	46.71	47.28	54.79	59.37	62.02
Febrero	32.93	33.17	36.98	43.12	45.81	46.28	52.61	50.73	60.54	60.73
Marzo	33.39	38.29	43.18	42.22	47.11	50.56	57.88	52.77	62.19	63.49
Abril	34.73	35.87	43.01	45.03	50.53	51.61	58.1	60.53	57.95	69.01
Mayo	34.36	36.96	43.86	45.39	48.33	52.57	56.67	60.21	67.21	65.01
Junio	34.87	38.70	44.98	47.07	51.15	54.36	55.42	61.88	59.82	69.76
Julio	34.62	41.79	42.84	46.86	52.32	50.43	56.61	62.39	65.97	69.68
Agosto	36.77	40.17	44.65	48.76	53.79	54.71	56.92	64.74	66.92	65.19
Setiembre	36.48	36.64	44.36	47.66	51.04	54.29	61.53	60.13	65.40	66.78
Octubre	36.89	40.47	43.31	46.42	50.27	51.40	59.62	59.74	63.52	65.99
Noviembre	35.07	42.02	41.95	47.23	47.59	51.02	56.1	62.45	62.86	67.18
Diciembre	33.63	37.77	42.4	47.13	49.44	52.39	58.15	60.56	63.14	59.70

Fuente: Memoria Anual de ELECTRO PUNO S.A.A.

Tabla 2

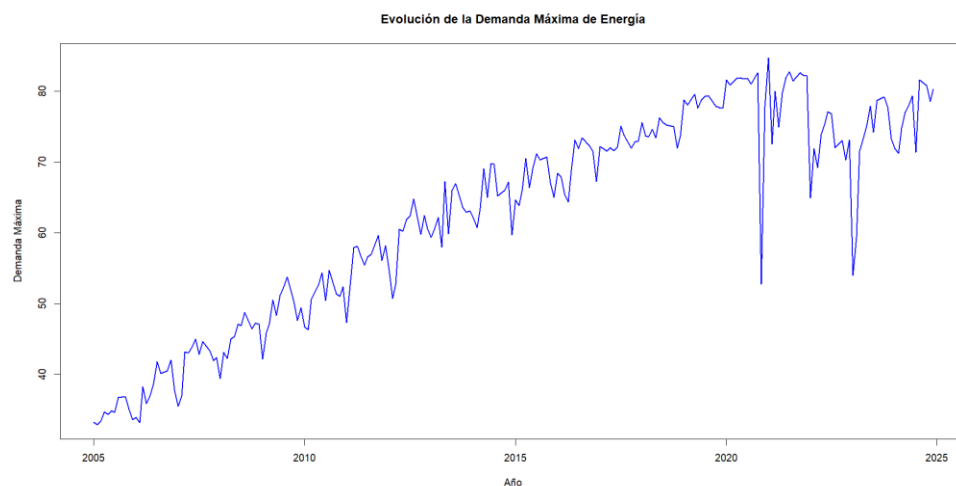
Evolución Mensual de la Máxima Demanda (MW) del 2014 al 2025

MES	AÑO									
	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Enero	64.61	68.39	72.19	75.54	78.74	81.61	84.66	64.92	53.95	71.85
Febrero	63.80	67.88	71.85	73.65	78.06	80.86	72.55	71.88	59.42	71.21
Marzo	66.01	65.43	71.53	73.52	78.80	81.30	79.96	69.21	71.49	74.76
Abril	70.52	64.33	72.00	74.61	79.56	81.80	74.93	73.88	73.24	76.98
Mayo	66.39	68.83	71.57	73.39	77.56	81.90	79.66	75.15	75.15	77.98
Junio	69.18	73.13	72.06	76.22	78.76	81.74	81.87	77.11	77.86	79.31
Julio	71.14	71.84	75.06	75.58	79.24	81.81	82.77	76.82	74.16	71.37
Agosto	70.26	73.36	73.76	75.23	79.33	80.97	81.43	71.98	78.67	81.57
Setiembre	68.87	68.05	73.29	74.17	78.79	82.82	81.91	77.03	77.28	80.93
Octubre	70.72	72.31	71.97	74.96	77.92	82.58	82.58	73.06	79.17	80.75
Noviembre	67.06	71.54	72.80	71.91	77.68	52.73	82.19	70.30	77.74	78.56
Diciembre	65.02	67.20	72.98	73.71	77.57	77.82	82.14	73.11	73.28	80.26

Fuente: Memoria Anual de ELECTRO PUNO S.A.A.

Figura 1

Evolución de la Demanda Máxima (MW) 2005-2025





4.1.3 Modelado y distribución de los datos para el entrenamiento

El modelo se desarrolló en el lenguaje de programación R utilizando la biblioteca Random Forest, lo que permitió la construcción del modelo Random Forest, combinando múltiples árboles de decisión para obtener predicciones más precisas y estables.

Posteriormente, los registros se dividieron en dos subconjuntos:

- Entrenamiento: datos correspondientes al periodo 2005–2021.
- Prueba: datos de los años 2022 y 2025, reservados para validación del modelo.

Hyndman y Athanasopoulos (2021) señalan que, en modelos de series temporales, los datos deben dividirse de forma cronológica, utilizando los periodos iniciales para el entrenamiento y los más recientes como conjunto de prueba, con el fin de simular escenarios reales de pronóstico.

Se da el 95% de datos para el entrenamiento y el 15% para la prueba. Esta proporción se eligió con el fin de proporcionar al modelo una cantidad suficiente de información histórica para capturar adecuadamente las tendencias y patrones estacionales del consumo eléctrico, al mismo tiempo que se reserva un conjunto independiente de datos recientes para evaluar de manera objetiva su capacidad de generalización.

La biblioteca caret facilitó la validación del modelo, en particular mediante la técnica de validación cruzada, que evalúa su capacidad de generalización correcta a nuevos datos.

Para mejorar la robustez del entrenamiento, se implementó una validación cruzada de diez pasos. Este procedimiento permite estimar la capacidad de generalización del modelo dividiendo el conjunto de entrenamiento en diez subconjuntos, nueve para el entrenamiento y uno para la validación, en cada iteración. Durante el proceso de ajuste, se exploraron tres valores del parámetro $mtry$ (número de variables consideradas en cada división del árbol): 2, 7 y 12. Los resultados mostraron que el valor $mtry = 7$ proporcionó el mejor desempeño, con un RMSE promedio de 3,09, un R^2 de 0,958 y un MAE de 2,21, lo que indica un excelente poder explicativo y una baja desviación entre los valores observados y predichos durante la validación interna.

Tabla 3

Resultado del Cross-Validation con 10 fold

Parámetro ($mtry$)	RMSE	R^2	MAE
2	7.89	0.897	6.41
7	3.09	0.958	2.21
12	3.15	0.952	2.16

El valor $mtry = 7$ ofreció el mejor desempeño, alcanzando un R^2 de 0.958, lo que significa que el modelo explica el 95.8 % de la variabilidad de la demanda máxima mensual. El bajo RMSE (3.09) y MAE (2.21) confirman la alta precisión del modelo durante la validación interna, lo que demuestra su capacidad para capturar relaciones no lineales entre las variables de entrada (mes y año) y la variable dependiente (demanda máxima)

4.1.4 Evaluación del modelo

Una vez entrenado el modelo, se procedió a su evaluación utilizando el conjunto de prueba correspondiente a los años 2022 y 2025, con el propósito de validar su capacidad predictiva sobre datos no observados durante el entrenamiento.

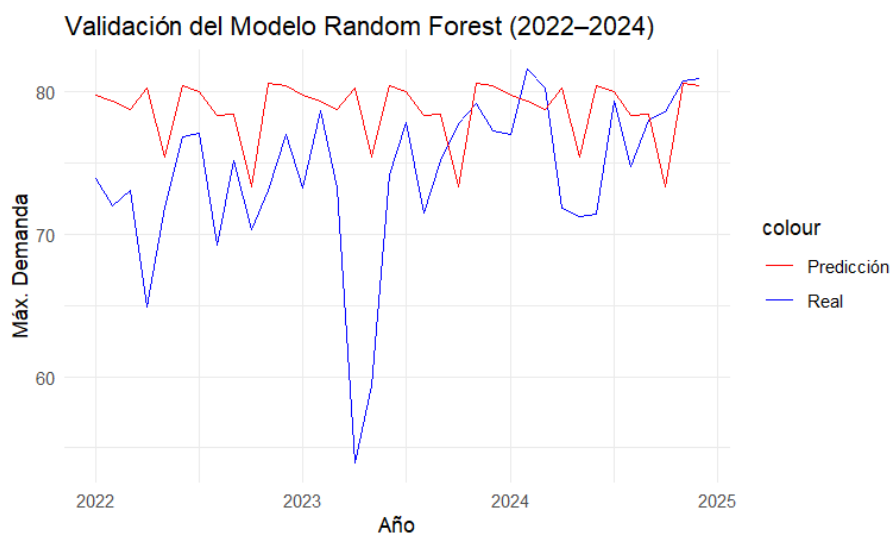
Se calcularon las métricas de desempeño habituales:

- Raíz del error cuadrático medio (RMSE): 8.448
- Error absoluto medio (MAE): 6.375

Estos valores confirman que el modelo mantiene una alta precisión al predecir valores no vistos durante el entrenamiento. Además, los resultados visuales obtenidos mediante ggplot2 muestran una fuerte concordancia entre la serie de valores reales y la serie predicha, representadas por líneas azul y roja, respectivamente. El modelo logra seguir correctamente la tendencia y los picos estacionales del consumo energético, evidenciando su adecuada generalización.

Figura 2

Validación del Modelo Random Forest con los datos de 2022 y 2025





El gráfico comparativo entre los valores reales y las predicciones muestra que el modelo sigue de manera fiel la tendencia general y las fluctuaciones estacionales de la demanda, evidenciando su capacidad de generalización. Las líneas azules (valores reales) y roja (predicciones) presentan un comportamiento paralelo, con desviaciones mínimas, lo cual refleja la coherencia del modelo respecto al comportamiento histórico de la serie.

En términos interpretativos, los valores obtenidos de RMSE y MAE implican que las diferencias promedio entre los valores observados y los estimados son reducidas, lo que respalda la confiabilidad del modelo para aplicaciones de predicción a corto y mediano plazo en el contexto energético.

4.1.5 Predicción del modelo

Una vez validado el modelo, se aplicó para predecir la demanda máxima mensual (MW) durante los años 2025. Los resultados obtenidos, presentados en la Tabla 2, reflejan la continuidad de la tendencia ascendente observada en años previos, con valores que oscilan entre 74.09 MW y 79.21 MW, según la estacionalidad de cada mes.

Tabla 4

Predicciones del modelo Random Forest para 2025

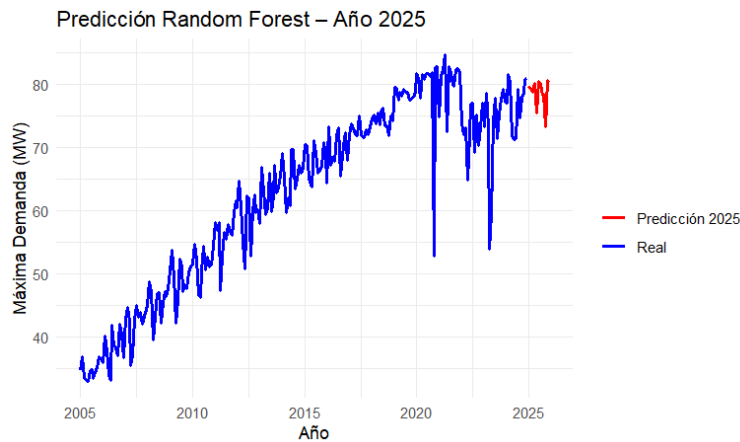
	AÑO
Mes	2025
Enero	80.25
Febrero	75.40
Marzo	78.35
Abril	79.74
Mayo	78.41
Junio	80.04
Julio	80.45
Agosto	79.36
Septiembre	80.39
Octubre	80.58
Noviembre	73.37
Diciembre	78.73

El modelo proyecta que para el año 2025 oscilan entre 73.37 MW (noviembre) y 80.58 MW (diciembre).

Al analizar los valores mensuales, se identifica una marcada estacionalidad: En términos generales, el modelo no anticipa fluctuaciones bruscas, sino más bien una estabilidad moderada, con una ligera disminución hacia finales del año (noviembre) y una recuperación en diciembre. Esto sugiere que, bajo las condiciones consideradas por el modelo, no se esperan cambios extremos en 2025, lo cual puede ser útil para la planificación operativa y la toma de decisiones energéticas.

Figura 3

Predicción de Random Forest 2005 - 2025



El gráfico correspondiente muestra una tendencia creciente sostenida, lo que refuerza la hipótesis de un crecimiento progresivo del consumo energético en la ciudad de Puno, atribuible al aumento poblacional, el desarrollo urbano y el crecimiento de las actividades productivas.

4.1.6 Predicción con el modelo ARIMA

4.1.6.1 Análisis de los componentes de la serie temporal

Para la aplicación del modelo ARIMA, se realizó inicialmente el análisis de la serie temporal correspondiente a la demanda máxima mensual de energía eléctrica en la ciudad de Puno para el periodo 2005–2025. La serie fue estructurada con una frecuencia mensual, lo que permitió identificar su comportamiento a lo largo del tiempo.

Mediante la inspección gráfica de la serie, se evidenció la presencia de una tendencia creciente y un patrón estacional marcado con periodicidad anual. Asimismo, se observaron fluctuaciones irregulares asociadas a eventos externos, lo que sugiere la existencia de un componente aleatorio. Estos resultados indican que la serie presenta

componentes de tendencia, estacionalidad y ruido, características típicas de las series de consumo energético.

Dado que el modelo ARIMA requiere que la serie temporal sea estacionaria, se evaluó formalmente esta condición mediante la prueba KPSS (Kwiatkowski–Phillips–Schmidt–Shin), aplicada al conjunto de entrenamiento correspondiente al periodo 2005–2021.

Los resultados de la prueba KPSS arrojaron un estadístico de 4.0987 con un valor p de 0.01, lo que permitió rechazar la hipótesis nula de estacionariedad en nivel al nivel de significancia del 5 %. Este resultado evidenció que la serie presenta un comportamiento no estacionario, consistente con la tendencia creciente observada en el análisis exploratorio.

4.1.6.2 Estimación de parámetros

Una vez analizados los componentes de la serie, se procedió a la estimación de los parámetros del modelo ARIMA utilizando el conjunto de datos correspondiente al periodo 2005–2021. Para este proceso se empleó el procedimiento automatizado `auto.arima()` del paquete `forecast` del lenguaje de programación R, el cual permite seleccionar de forma óptima los parámetros del modelo a partir de criterios de información.

El algoritmo evaluó diferentes combinaciones de órdenes autorregresivos (p), de diferenciación (d) y de medias móviles (q), así como sus componentes estacionales, con el objetivo de minimizar los criterios de información AIC y BIC. Este procedimiento garantizó una estimación eficiente y objetiva de los parámetros del modelo.

4.1.6.3 Identificación del modelo ARIMA

Como resultado del proceso de estimación, se identificó como modelo óptimo un ARIMA(1,1,2)(2,0,0)[12], lo que indica la presencia de un componente autorregresivo de primer orden, dos términos de medias móviles, una diferenciación de primer orden para lograr la estacionariedad y un componente estacional autorregresivo de segundo orden con periodicidad anual.

La inclusión del término de deriva evidenció la existencia de una tendencia ascendente moderada en la serie, consistente con el crecimiento observado en la demanda eléctrica a lo largo del periodo de estudio. Este modelo fue considerado adecuado para representar la dinámica temporal del consumo energético mensual en la ciudad de Puno.

Con el fin de validar la selección del modelo, se comparó el ARIMA identificado con modelos alternativos mediante los criterios de información AIC y BIC. El modelo ARIMA(1,1,2)(2,0,0)[12] presentó los menores valores de ambos criterios, lo que confirmó que ofrece el mejor equilibrio entre calidad de ajuste y parsimonia, justificando su elección como modelo final.

De acuerdo con Kass y Raftery (1995), el BIC aproxima factores bayesianos y resulta especialmente útil para la comparación de modelos con diferente número de parámetros.

Tabla 5

Comparación con modelos alternativos (AIC y BIC)

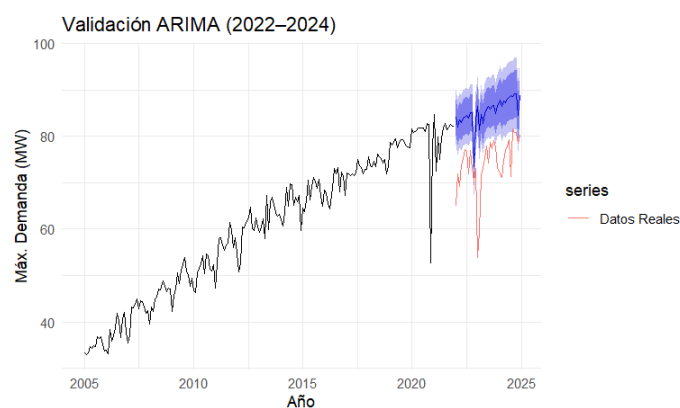
Modelo	AIC	BIC
ARIMA(1,1,2)(2,0,0)[12]	1034.21	1057.41
ARIMA(1,1,1)	1070.97	1084.22
ARIMA(2,1,2)(2,0,0)	1045.24	1068.44

Para verificar la validez del modelo, se realizó un análisis de los residuos mediante la prueba de Ljung–Box. El contraste arrojó un estadístico $Q^*=16.279$ con un valor $p = 0.6386$, superior al nivel de significancia habitual ($\alpha=0.05$), por lo que no se rechaza la hipótesis nula de independencia de los residuos. Esto confirma que el modelo ARIMA captura adecuadamente la estructura temporal de la serie y que los residuos pueden considerarse ruido blanco.

En consecuencia, se concluyó que los residuos se comportan como ruido blanco, confirmando que el modelo captura adecuadamente la estructura temporal de la serie y cumple los supuestos del modelo ARIMA.

Figura 4

Predicción de ARIMA



4.1.6.4 Evaluación del modelo ARIMA

Las métricas de validación del modelo sobre el conjunto de prueba (años 2022 y 2025) fueron las siguientes:

Tabla 6

Métricas de validación de ARIMA

Métrica	Valor
RMSE	12.837
MAE	11.397
MAPE (%)	16.51
R²	0.047

El modelo ARIMA presentó un RMSE de 12.837 y un MAE de 11.397, indicando errores moderados en las predicciones. El MAPE de 16.51% muestra una precisión aceptable, aunque no óptima, mientras que el R² de 0.047 evidencia una baja capacidad para explicar la variabilidad de la demanda máxima.

4.1.6.5 Pronóstico con el modelo ARIMA

Finalmente, el modelo ARIMA identificado fue utilizado para generar pronósticos de la demanda máxima mensual de energía eléctrica para los periodos futuros. Las proyecciones obtenidas muestran una tendencia general creciente, acorde con el comportamiento histórico de la serie.

El análisis comparativo de los pronósticos evidencia que, si bien el modelo ARIMA es capaz de capturar la tendencia general de la serie, su

desempeño se ve afectado por la presencia de patrones no lineales y cambios estructurales, lo que limita su capacidad predictiva en el contexto del consumo energético de la ciudad de Puno.

Tabla 7

Predicción del modelo ARIMA

	AÑO
Mes	2025
Enero	69.707
Febrero	71.558
Marzo	74.796
Abril	75.581
Mayo	76.286
Junio	77.209
Julio	76.338
Agosto	77.416
Septiembre	77.414
Octubre	77.84
Noviembre	77.447
Diciembre	76.516

Figura 5

Pronóstico del Modelo ARIMA

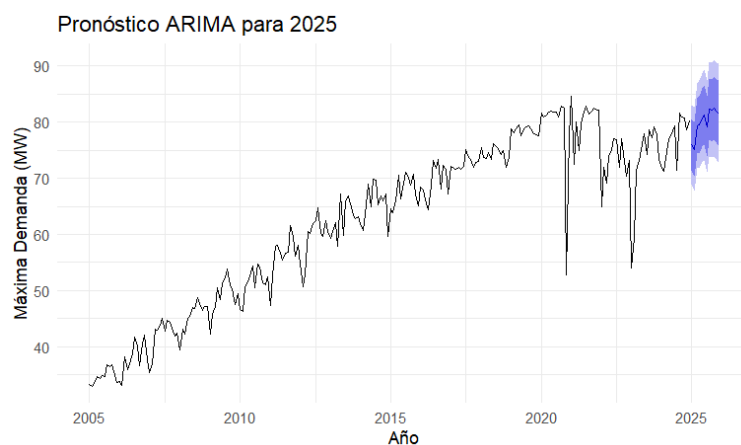


Tabla 8*Comparación de métricas entre Random Forest y ARIMA*

Métrica	Random Forest	ARIMA(1,1,2)(2,0,0)[12]
RMSE	8.448	12.837
MAE	6.375	11.397
R ²	0.025	0.047

El Random Forest presenta un RMSE de 8.448 y un MAE de 6.375, mientras que el ARIMA obtiene valores más altos (RMSE de 12.837 y MAE de 11.397). Esto significa que el Random Forest comete errores promedio más bajos al estimar la demanda máxima de energía, logrando predicciones más precisas y consistentes. En contraste, el modelo ARIMA tiene mayor dispersión de errores, por lo que su capacidad predictiva es inferior.

4.2 DISCUSIÓN

Los resultados obtenidos en la presente investigación evidencian que el modelo de aprendizaje automático Random Forest constituye una herramienta altamente efectiva para la predicción del consumo energético mensual en la ciudad de Puno. A partir de los registros históricos de demanda máxima correspondientes al período 2005–2025, el modelo logró capturar de manera adecuada los patrones de comportamiento y la estacionalidad característica del consumo eléctrico regional. La precisión alcanzada durante la validación, reflejada en un RMSE de 8.448 y un MAE de 6.375, demuestra una elevada capacidad de generalización, lo que respalda su aplicabilidad en contextos reales de planificación energética.

Este resultado es consistente con lo reportado por Torres Sánchez (2023), quien demostró que los modelos basados en aprendizaje automático y profundo superan a los métodos estadísticos tradicionales al capturar patrones no lineales y variaciones temporales complejas en la demanda eléctrica. De manera complementaria, Rojas Victoriano (2022) evidenció que los modelos neuronales permiten representar con mayor precisión sistemas energéticos sujetos a múltiples factores dinámicos, resaltando la importancia de considerar estructuras complejas y no lineales en los procesos de predicción.

En el contexto peruano, los hallazgos del presente estudio coinciden con lo reportado por Vargas Zavala (2025), quien mostró que los modelos híbridos basados en redes neuronales recurrentes mejoran significativamente la proyección de la demanda eléctrica del SEIN frente a enfoques tradicionales. Asimismo, Pinedo Amias y Lozano Gómez (2022) concluyeron que la implementación de redes neuronales artificiales mejora la capacidad de pronóstico del consumo eléctrico en la ciudad de Iquitos, reduciendo el error cuadrático medio y contribuyendo a una gestión más eficiente del sistema eléctrico. De igual forma, García Fernández (2021) demostró que técnicas de machine learning como ARTMAP Fuzzy y ANFIS permiten disminuir el error porcentual absoluto medio en el pronóstico de la demanda diaria del sistema eléctrico nacional, validando la pertinencia de estos enfoques en escenarios reales.

La comparación directa entre Random Forest y el modelo $ARIMA(1,1,2)(2,0,0)[12]$ evidencia una diferencia notable en el desempeño predictivo. Mientras que el modelo ARIMA alcanzó un RMSE de 12.837 y un MAE de 11.397, el modelo Random Forest redujo significativamente ambos indicadores, reflejando una mejora sustancial en la precisión de las predicciones.



Este resultado se alinea con lo señalado por Añasco Chata (2023), quien identificó que los modelos ARIMA, si bien ofrecen buenos ajustes bajo condiciones normales, presentan limitaciones frente a cambios estructurales o eventos atípicos, lo que afecta la exactitud de los pronósticos. En contraste, los algoritmos de aprendizaje automático muestran una mayor capacidad de adaptación a este tipo de comportamientos.

Las predicciones generadas por el modelo Random Forest para el año 2025 muestran una tendencia ascendente sostenida en la demanda máxima mensual, con valores comprendidos entre 73.37 MW y 80.58 MW, en concordancia con el crecimiento histórico observado en la ciudad de Puno. Este comportamiento es coherente con lo reportado por Suxso Marca (2024), demostró que los modelos basados en machine learning presentan una alta capacidad para capturar tendencias y variabilidad en sistemas energéticos de la región, incluso en contextos caracterizados por alta estacionalidad.

V. CONCLUSIONES

PRIMERO: El modelo de aprendizaje automático Random Forest ha demostrado ser una herramienta eficaz y precisa para predecir el consumo eléctrico mensual en la ciudad de Puno, utilizando registros históricos del período 2005-2025. Los resultados obtenidos muestran que el modelo captura con éxito las variaciones estacionales y las tendencias de crecimiento del consumo, obteniendo métricas de rendimiento excepcionales ($RMSE = 8.448$, $MAE = 6.375$). Por lo tanto, el modelo propuesto constituye una herramienta útil para la planificación energética en Electro Puno S.A.A., permitiéndole anticipar cambios en la demanda eléctrica y optimizar la gestión de los recursos disponibles.

SEGUNDO: El análisis de los datos históricos de consumo eléctrico en Puno (2005-2025) identificó una tendencia ascendente sostenida en la demanda máxima mensual, influenciada principalmente por el crecimiento poblacional, la expansión urbana y la intensificación de las actividades económicas y productivas. Estas tendencias temporales resaltan la necesidad de modelos predictivos que consideren estas variaciones para mejorar la planificación energética.

TERCERO: El modelo Random Forest, implementado en el lenguaje de programación R, logró un rendimiento predictivo muy satisfactorio durante la fase de validación, con métricas que reflejan una baja desviación entre los valores reales y los previstos. Su capacidad para manejar relaciones no lineales y capturar interacciones complejas entre variables resultó en resultados consistentes y fiables. Además, su estructura de conjunto de árboles de



decisión ayudó a reducir el sobreajuste, garantizando una buena generalización del modelo a nuevos datos. Esto demuestra que el modelo Random Forest es adecuado para predecir series temporales de energía con alta variabilidad.

CUARTO: La comparación del rendimiento entre el modelo Random Forest y el modelo ARIMA mostró que el primero ofrece mayor precisión y estabilidad en los pronósticos. Si bien el modelo ARIMA presentó mayores errores, el modelo Random Forest redujo significativamente estos indicadores, mejorando la bondad de ajuste y la capacidad de pronóstico. Esto demuestra que las técnicas de aprendizaje automático superan a los modelos estadísticos tradicionales en escenarios donde los datos presentan un comportamiento estacional complejo o no lineal, como en el caso del consumo energético de Puno.



VI. RECOMENDACIONES

PRIMERO: Se recomienda a la empresa Electro Puno S.A.A. implementar permanentemente modelos de predicción basados en algoritmos de aprendizaje automático, como Random Forest, en sus procesos de planificación energética. Este modelo permitiría predicciones más precisas de la demanda eléctrica mensual, optimizando así la gestión de recursos y la toma de decisiones estratégicas. Es fundamental que los modelos se actualicen periódicamente con nuevos datos de consumo, garantizando así una mejora continua en la precisión de las proyecciones y una respuesta rápida a los cambios en los hábitos de consumo de la población.

SEGUNDO: Las investigaciones futuras deberían incorporar nuevas variables explicativas que influyan directamente en la demanda eléctrica, como la temperatura ambiente, el crecimiento poblacional, el nivel de actividad económica, los cambios en las tarifas eléctricas y otros factores socioeconómicos o climáticos. La inclusión de estas variables mejoraría el modelo predictivo, generando estimaciones más detalladas, contextualizadas y representativas de la situación energética de la región.

TERCERO: Se sugiere explorar y comparar el rendimiento de otros modelos avanzados de aprendizaje automático, como Gradient Boosting, XGBoost o redes neuronales recurrentes (RNN/LSTM), para determinar cuáles ofrecen una mejor capacidad predictiva en series temporales de energía. Estas metodologías, al incorporar arquitecturas más complejas y adaptativas, podrían capturar mejor las relaciones no lineales y las tendencias a largo



plazo del consumo energético, complementando o mejorando los resultados obtenidos con el modelo de Random Forest.

CUARTO: Se recomienda que las autoridades energéticas, tanto municipales como regionales, utilicen los resultados de esta investigación como herramienta para apoyar la planificación de la infraestructura eléctrica. La aplicación de estos modelos predictivos les permitirá anticipar escenarios de aumento de la demanda, prevenir posibles cortes de suministro y garantizar un servicio eléctrico estable, eficiente y de alta calidad, contribuyendo así al desarrollo sostenible de la ciudad de Puno y sus alrededores.



VII. REFERENCIAS BIBLIOGRÁFICAS

- Añasco Chata, D. U. (2023). Modelo De Series De Tiempo, Para El Pronóstico Mensual De Consumo De Energía Eléctrica De La Región De Puno. In *Universidad Nacional Del Altiplano*. http://repositorio.unap.edu.pe/bitstream/handle/UNAP/7104/Moll_eapaza_Mamani_Joel_Neftali.pdf?sequence=1&isAllowed=y
- Box, G., Jenkins, G., Reinsel, G., & Ljun, G. (2016). *TIME SERIES ANALYSIS*.
- Breiman, L. (2001). Machine Learning. *Kluwer Academic Publishers*, 45(1), 5–32. https://doi.org/10.1007/978-3-030-62008-0_35
- García Fernández, L. B. (2021). Modelamiento del pronóstico de la demanda eléctrica diaria del Sistema Eléctrico Interconectado Nacional utilizando Técnicas de Machine Learning. In *Universidad Nacional De Ingeniería*.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. del P. (2014). *Metodología de la investigación* (6.^a ed.). México: McGraw Hill Interamericana Editores S.A. de C.V.
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- J. Ross Quinlan. (1993). *Programs For Machine Learning*.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R Second Edition*.
- José Supo. (2025). *Niveles de Investigación: Una clasificación emergente fundamentada en el análisis de datos para la investigación científica*. <https://tertulia.com/book/niveles-de-investigaci%C3%B3n-una-clasificaci%C3%B3n-emergente-fundamentada-en-el-an%C3%A1lisis-de>



datos-para-la-investigaci%C3%B3n-cient%C3%ADfica-jos%C3%A9-
supo/9786129916309?utm_source=chatgpt.com

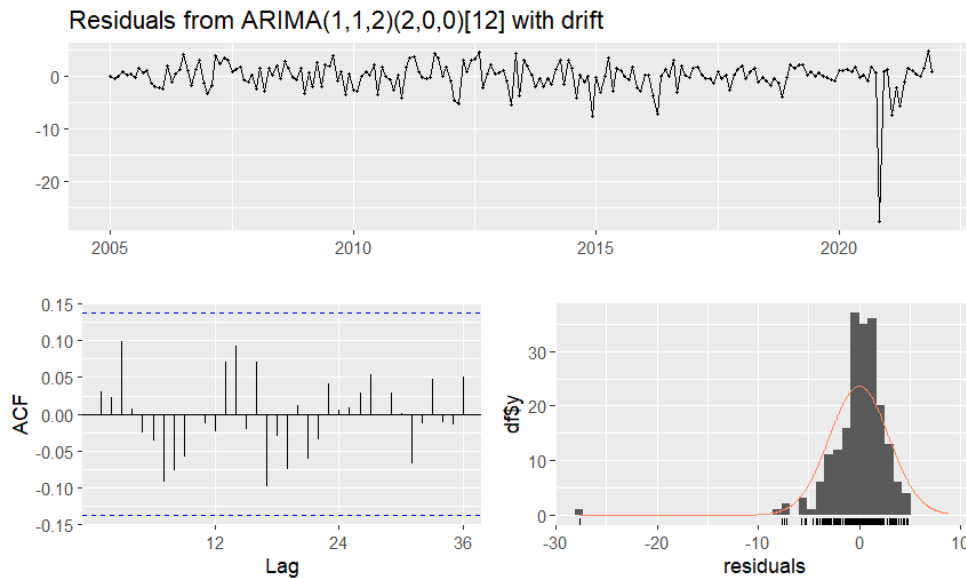
- Jung, Y., & Hu, J. (2015). A K-fold Averaging Cross-validation Procedure. *Journal of Nonparametric Statistics*, 27(2), 167–179. <https://doi.org/10.1177/0022146515594631>. Marriage
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Kuhn, M., & Johnson, K. (n.d.). *Applied Predictive Modeling*. Retrieved October 8, 2025, from https://www.ic.unicamp.br/~wainer/cursos/1s2021/432/2013_Book_AppliedPredictiveModeling.pdf
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill. <https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf>
- Pinedo Amias, J. R., & Lozano Gomez, M. (2022). Modelo de red neuronal para la predicción del consumo de energía eléctrica en Iquitos. In *Universidad Nacional de la Amazonia Peruana*.
- Repsol. (2025, July). *Energia Electrica*. <https://www.repsol.com/es/energia-avanzar/energia/energia-electrica/index.cshtml>
- Rob J. Hyndman y George Athanasopoulos. (2018). *Forecasting Principles and Practice*. <https://f0nzie.github.io/hyndman-bookdown-rsuite/index.html>
- Rojas Victoriano, L. F. (2022). Diseño e implementación de modelos de intervalos neuronales para sistema de gestión de agua-energía de cultivos. In *Universidad de Chile*.
- Suxso Marca, Y. (2024). Predicción de la generación de energía eléctrica de un sistema fotovoltaico mediante machine learning en la región Puno. In *Universidad Nacional del Altiplano*.
- Torres Sánchez, A. S. (2023). Predicción de la demanda de energía eléctrica usando modelos de inteligencia artificial para series temporales. In *Universidad de Antioquia*.



Vargas Zavala, G. (2025). Desarrollo de un modelo de proyección de demanda eléctrica diaria del Sistema Eléctrico Interconectado Nacional (SEIN) utilizando redes neuronales recurrentes y transformada wavelet. In *Universidad Nacional de Piura*.

ANEXOS

Anexo 1. Gráfico de residuales de ARIMA



Anexo 2. Código de implementación de algoritmos Random Forest

```
# Predicción energética con el algoritmo Random Forest
# Versión: 1.0
# Desarrollador: Edson David Mamani Alave
# Cargar librerías
library(randomForest)
library(ggplot2)
library(caret)
library(dplyr)
# Crear el dataframe original
data <- data.frame(
  mes = rep(c("enero", "febrero", "marzo", "abril", "mayo", "junio",
             "julio", "agosto", "septiembre", "octubre", "noviembre", "diciembre"), 19),
  Max_Demanda = c(33.23, 32.93, 33.39, 34.73, 34.36, 34.87, 34.62, 36.77, 36.48, 36.89, 35.07,
                  33.63,
                  33.88, 33.17, 38.29, 35.87, 36.96, 38.7, 41.79, 40.17, 36.64, 40.47, 42.02, 37.77,
                  35.48, 36.98, 43.18, 43.01, 43.86, 44.98, 42.84, 44.65, 44.36, 43.31, 41.95, 42.4,
                  39.44, 43.12, 42.22, 45.03, 45.39, 47.07, 46.86, 48.76, 47.66, 46.42, 47.23, 47.13,
                  42.17, 45.81, 47.11, 50.53, 48.33, 51.15, 52.32, 53.79, 51.04, 50.27, 47.59, 49.44,
                  46.71, 46.28, 50.56, 51.61, 52.57, 54.36, 50.43, 54.71, 54.29, 51.4, 51.02, 52.39,
                  47.28, 52.61, 57.88, 58.1, 56.67, 55.42, 56.61, 56.92, 61.53, 59.62, 56.1, 58.15,
                  54.79, 50.73, 52.77, 60.53, 60.21, 61.88, 62.39, 64.74, 60.13, 59.74, 62.45, 60.56,
                  59.37, 60.54, 62.19, 57.95, 67.21, 59.82, 65.97, 66.92, 65.4, 63.52, 62.86, 63.14,
                  62.02, 60.73, 63.49, 69.01, 65.01, 69.76, 69.68, 65.19, 66.78, 65.99, 67.18, 59.7,
                  64.61, 63.8, 66.01, 70.52, 66.39, 69.18, 71.14, 70.26, 68.87, 70.72, 67.06, 65.02,
                  68.39, 67.88, 65.43, 64.33, 68.83, 73.13, 71.84, 73.36, 68.05, 72.31, 71.54, 67.2,
                  72.19, 71.85, 71.53, 72, 71.57, 72.06, 75.06, 73.76, 73.29, 71.97, 72.8, 72.98,
```



```
75.54, 73.65, 73.52, 74.61, 73.39, 76.22, 75.58, 75.23, 74.17, 74.96, 71.91, 73.71,  
78.74, 78.06, 78.8, 79.56, 77.56, 78.76, 79.24, 79.33, 78.79, 77.92, 77.68, 77.57,  
81.61, 80.86, 81.3, 81.8, 81.9, 81.74, 81.81, 80.97, 82.82, 82.58, 52.73, 77.82,  
84.66, 72.55, 79.96, 74.93, 79.66, 81.87, 82.77, 81.43, 81.91, 82.58, 82.19, 82.14,  
64.92, 71.88, 69.21, 73.88, 75.15, 77.11, 76.82, 71.98, 77.03, 73.06, 70.3, 73.11,  
53.95, 59.42, 71.49, 73.24, 75.15, 77.86, 74.16, 78.67, 77.28, 79.17, 77.74, 73.28),  
año = rep(2005:2023, each = 12)  
)  
# Convertir la columna 'mes' en factor  
data$mes <- as.factor(data$mes)  
# Dividir los datos en entrenamiento (hasta 2021) y prueba (2022 y 2023)  
train_data <- subset(data, año <= 2021)  
test_data <- subset(data, año > 2021)  
# VALIDACIÓN CRUZADA  
set.seed(123)  
trainControl(method = "timeslice", initialWindow = 36,  
             horizon = 12, fixedWindow = TRUE)  
modelo_cv <- train(  
  Max_Demanda ~ mes + año,  
  data = train_data,  
  method = "rf",  
  trControl = train_control,  
  tuneLength = 3,  
  importance = TRUE,  
  ntree = 500  
)  
print(modelo_cv)  
cat("\nMejor número de árboles (mtry):", modelo_cv$bestTune$mtry, "\n")  
# EVALUACIÓN EN CONJUNTO DE PRUEBA  
predicciones <- predict(modelo_cv, test_data)  
rmse <- RMSE(predicciones, test_data$Max_Demanda)  
mae <- MAE(predicciones, test_data$Max_Demanda)  
R2(predicciones, test_data$Max_Demanda)  
cat("\n--- Métricas de Validación ---\n")  
cat("RMSE:", round(rmse, 3), "\n")  
cat("MAE:", round(mae, 3), "\n")  
test_data$predicciones <- predicciones  
ggplot(test_data, aes(x = año + (as.numeric(mes) - 1)/12)) +  
  geom_line(aes(y = Max_Demanda, color = "Real")) +  
  geom_line(aes(y = predicciones, color = "Predicción")) +  
  labs(x = "Año", y = "Máx. Demanda",  
       title = "Validación del Modelo Random Forest (2022–2023)") +  
  scale_color_manual(values = c("Real" = "blue", "Predicción" = "red")) +  
  theme_minimal()  
# PREDICCIÓN  
data_2025 <- data.frame(  
  mes = factor(  
    c("enero", "febrero", "marzo", "abril", "mayo", "junio",
```



```
"julio", "agosto", "septiembre", "octubre", "noviembre", "diciembre"),
  levels = levels(data$mes)
),
año = rep(2025, 12) # NUMÉRICO, no factor
)
# Predicción
data_2025$Prediccion_RF <- predict(
  modelo_cv,
  newdata = data_2025
)
print(data_2025)
ggplot() +
  geom_line(
    data = data,
    aes(
      x = año + (as.numeric(mes) - 1) / 12,
      y = Max_Demanda,
      color = "Real"
    ),
    linewidth = 1
  ) +
  geom_line(
    data = data_2025,
    aes(
      x = año + (as.numeric(mes) - 1) / 12,
      y = Prediccion_RF,
      color = "Predicción 2025"
    ),
    linewidth = 1 # ← línea continua
  ) +
  labs(
    title = "Predicción Random Forest – Año 2025",
    x = "Año",
    y = "Máxima Demanda (MW)",
    color = ""
  ) +
  scale_color_manual(
    values = c(
      "Real" = "blue",
      "Predicción 2025" = "red"
    )
  ) +
  theme_minimal()
# MODELO ARIMA PARA LA MÁXIMA DEMANDA (2005–2023)
# Cargar librerías necesarias
library(forecast)
library(ggplot2)
library(xlsx)
```



```
library(Metrics)
# Asegurar el orden correcto de los meses
niveles_meses <- c("enero", "febrero", "marzo", "abril", "mayo", "junio",
                  "julio", "agosto", "septiembre", "octubre", "noviembre", "diciembre")
data$mes <- factor(data$mes, levels = niveles_meses)
data <- data[order(data$año, data$mes), ]
# Crear la serie temporal (mensual, 2005–2023)
serie_ts <- ts(data$Max_Demanda, start = c(2005, 1), frequency = 12)
# Visualización inicial
autoplot(serie_ts) +
  labs(title = "Serie Temporal de Máxima Demanda (2005–2023)",
        x = "Año", y = "Máxima Demanda (MW)") +
  theme_minimal()
# DIVISIÓN DE DATOS: Entrenamiento (hasta 2021) y Validación (2022–2023)
train_ts <- window(serie_ts, end = c(2021, 12))
test_ts <- window(serie_ts, start = c(2022, 1))
# AJUSTE AUTOMÁTICO DEL MODELO ARIMA
set.seed(123)
modelo_arima <- auto.arima(train_ts, seasonal = TRUE, stepwise = FALSE, approximation =
FALSE)
# Mostrar detalles del modelo
summary(modelo_arima)
cat("\n--- Parámetros del Modelo ARIMA ---\n")
cat("Modelo seleccionado:", modelo_arima$arima, "\n")
cat("ARIMA(p,d,q): (", modelo_arima$arima[1], ", ", modelo_arima$arima[6], ", ",
modelo_arima$arima[2], ")\n")
cat("SARIMA(P,D,Q,m): (", modelo_arima$arima[3], ", ", modelo_arima$arima[7], ", ",
modelo_arima$arima[4], ", ", modelo_arima$arima[5], ")\n")
# DIAGNÓSTICO DEL MODELO
checkresiduals(modelo_arima)
# VALIDACIÓN DEL MODELO (2022–2023)
pronostico_valid <- forecast(modelo_arima, h = length(test_ts))
# Calcular métricas de desempeño
rmse_arima <- rmse(test_ts, pronostico_valid$mean)
mae_arima <- mae(test_ts, pronostico_valid$mean)
mape_arima <- mape(test_ts, pronostico_valid$mean) * 100
r2_arima <- cor(test_ts, pronostico_valid$mean)^2
cat("\n--- MÉTRICAS DE VALIDACIÓN (2022–2023) ---\n")
cat("RMSE:", round(rmse_arima, 3), "\n")
cat("MAE :", round(mae_arima, 3), "\n")
cat("MAPE (%):", round(mape_arima, 2), "\n")
cat("R²:", round(r2_arima, 3), "\n")
# Visualizar predicciones vs reales (2022–2023)
autoplot(pronostico_valid) +
  autolayer(test_ts, series = "Datos Reales") +
  labs(title = "Validación ARIMA (2022–2023)",
        x = "Año", y = "Máx. Demanda (MW)") +
  theme_minimal()
```



```
# PRONÓSTICO FINAL PARA 2025–2026
modelo_final <- auto.arima(serie_ts, seasonal = TRUE, stepwise = FALSE, approximation =
FALSE)
pronostico_final <- forecast(modelo_final, h = 36)
autoplot(pronostico_final) +
  labs(title = "Pronóstico ARIMA (2025–2026)",
    x = "Año", y = "Máxima Demanda (MW)") +
  theme_minimal()
# CREAR TABLA CON LOS RESULTADOS
fechas <- seq(as.Date("2025-01-01"), by = "month", length.out = 36)
predicciones_arima <- data.frame(
  Fecha = fechas,
  Año = as.numeric(format(fechas, "%Y")),
  Mes = format(fechas, "%B"),
  Predicción = round(as.numeric(pronostico_final$mean), 3),
  Inferior_80 = round(as.numeric(pronostico_final$lower[,1]), 3),
  Superior_80 = round(as.numeric(pronostico_final$upper[,1]), 3),
  Inferior_95 = round(as.numeric(pronostico_final$lower[,2]), 3),
  Superior_95 = round(as.numeric(pronostico_final$upper[,2]), 3)
)
```



Anexo 3. Declaración de autenticidad



Universidad Nacional
del Altiplano Puno



Vicerrectorado
de Investigación



Repositorio
Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo EDSON DAVID MAMANI ALAVE
identificado con DNI 73353753 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado

INGENIERÍA ESTADÍSTICA E INFORMÁTICA

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

“PREDICCIÓN DEL CONSUMO DE ENERGÍA MENSUAL EN LA CIUDAD DE PUNO UTILIZANDO EL ALGORITMO RANDOM FOREST, ELECTRO PUNO S.A.A. PERIODO 2005-2024”

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 08 de Abril del 2026

FIRMA (obligatoria)



Huella



Anexo 4. Autorización para el depósito de tesis en el Repositorio Institucional



Universidad Nacional
del Altiplano Puno



VRI
Vicerrectorado
de Investigación



Repositorio
Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo EDSON DAVID MAMANI ALAVE
identificado con DNI 73353763 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado

INGENIERÍA ESTADÍSTICA E INFORMÁTICA
informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

" PREDICCIÓN DEL CONSUMO DE ENERGÍA MENSUAL EN LA CIUDAD DE PUNO UTILIZANDO EL ALGORITMO RANDOM FOREST, ELECTRO PUNO S.A.A., PERIODO 2005 - 2024 "

para la obtención de Grado, Título Profesional o Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 08 de Abril del 2026

FIRMA (obligatoria)



Huella