

## **Maestría en**

# **CIENCIA DE DATOS Y MAQUINAS DE APRENDIZAJE CON MENCIÓN EN INTELIGENCIA ARTIFICIAL**

## **Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Maquinas de Aprendizaje con Mención en Inteligencia Artificial**

### **TEMA:**

**EVALUACIÓN Y COMPARATIVA DE ALGORITMOS DE MACHINE LEARNING PARA  
LA PREDICCIÓN DE LA DEMANDA ENERGÉTICA NACIONAL**

### **AUTORES:**

Eduardo Javier Amaya Oñate

Jorge Leonardo Vidal Zambrano

Fabricio Enrique Villavicencio Ramos

Christian Alexis Yugcha Alomaliza

### **AFILIACIÓN INSTITUCIONAL:**

Universidad Internacional del Ecuador - UIDE Quito

### **TUTOR:**

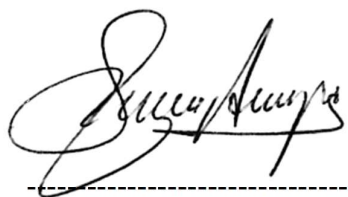
Karla Estefanía Mora Cajas

Fernanda Paulina Vizcaíno

## CERTIFICACIÓN DE AUTORÍA

Nosotros, **Eduardo Javier Amaya Oñate, Jorge Leonardo Vidal Zambrano, Fabricio Enrique Villavicencio Ramos, Christian Alexis Yugcha Alomaliza**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales



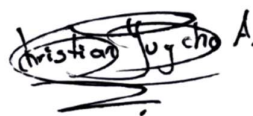
Firma  
**Eduardo Javier Amaya Oñate**



Firma  
**Jorge Leonardo Vidal Zambrano**



Firma  
**Fabricio Enrique Villavicencio Ramos**

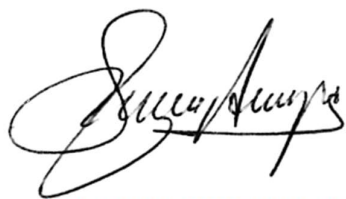


Firma  
**Christian Alexis Yugcha Alomaliza**

## AUTORIZACIÓN DE DERECHOS DE PROPIEDAD INTELECTUAL

Nosotros, **Eduardo Javier Amaya Oñate, Jorge Leonardo Vidal Zambrano, Fabricio Enrique Villavicencio Ramos, Christian Alexis Yugcha Alomaliza**, en calidad de autores del trabajo de investigación titulado *Evaluación y comparativa de algoritmos de machine learning para la predicción de la demanda energética nacional*, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, diciembre 2025



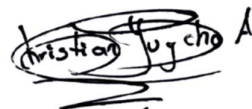
Firma  
**Eduardo Javier Amaya Oñate**



Firma  
**Jorge Leonardo Vidal Zambrano**



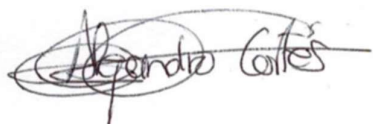
Firma  
**Fabricio Enrique Villavicencio Ramos**



Firma  
**Christian Alexis Yugcha Alomaliza**

## APROBACIÓN DE DIRECCIÓN Y COORDINACIÓN DEL PROGRAMA

Nosotros, **Alejandro Cortés López** Director EIG y **Karla Estefanía Mora Cajas** Coordinadora UIDE, declaramos que: **Eduardo Javier Amaya Oñate, Jorge Leonardo Vidal Zambrano, Fabricio Enrique Villavicencio Ramos, Christian Alexis Yugcha Alomaliza**, son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



-----  
**Alejandro Cortés López**

Director de la

Maestría en Ciencia de Datos y Maquinas  
de Aprendizaje con Mención en Inteligencia  
Artificial



-----  
**Karla Estefanía Mora Cajas**

Coordinadora de la

Maestría en Ciencia de Datos y Maquinas de  
Aprendizaje con Mención en Inteligencia  
Artificial

## **DEDICATORIA**

Dedicamos el siguiente trabajo de investigación a nuestros padres y seres queridos quienes siempre están apoyándonos y motivándonos para culminar nuestras metas académicas.

## **AGRADECIMIENTOS**

Agradecemos en primer lugar a nuestras familias y seres queridos quienes nos han apoyado desde siempre en nuestro proceso académico y profesional, así mismo a la universidad internacional del ecuador y al programa de la maestría de Ciencias de datos y máquinas de aprendizaje por los conocimientos brindados y la formación académica recibida.

Se agradece de manera especial a todos los docentes, la tutora Karla Estefanía Mora Cajas y al director de la maestría Alejandro Cortés López, por la guía, el apoyo, las observaciones y enseñanzas a lo largo de todo el proceso de estudio e investigación.

Por último, extendemos el agradecimiento a nuestros compañeros de maestría quienes en algún momento compartimos trabajos grupales y entregables. Gracias por su contribución y su apoyo en este camino de titulación.

## RESUMEN

La presente tesis comprende la evaluación comparativa de algoritmos de predicción aplicados a la demanda energética nacional en el Ecuador, para así, identificar un modelo que nos permita hacer un estimado del consumo eléctrico basándonos en información obtenida entre el 2022 y 2025.

Esta investigación usa la metodología KDD, la cual tiene un enfoque estructurado en descubrimiento de patrones en los datos integrando la preparación, transformación y modelado de series temporales. Como parte del desarrollo, se ha planteado un prototipo de aplicación denominado ELFO (Electric Forecast) que soluciona y automatiza el proceso de predicción, desde el procesamiento ETL hasta la generación de pronósticos.

La obtención de datos proviene de reportes oficiales de consumo, los cuales son archivos de tipo Excel con registros hora a hora. Estos datos requirieron consolidación y depuración para asegurar su consistencia. En la fase de transformación, se aplicaron procedimientos de detección de anomalías mediante Z-score, manejo de valores faltantes, imputación por interpolación lineal y, por último, una agregación temporal diaria que reduce el ruido de alta frecuencia y facilita el aprendizaje de tendencias estacionales.

En el modelado se usó un enfoque comparativo entre dos arquitecturas predictivas. El modelo estadístico aditivo basado en Facebook Prophet y un modelo de Machine Learning supervisado mediante XGBoost (usando ingeniería de características temporales y optimización de hiperparámetros mediante búsqueda bayesiana).

La evaluación del desempeño se realizó usando métricas estándar como MAE, RMSE y MAPE, complementadas con visualizaciones comparativas entre los valores reales y los predichos.

Los resultados que se obtuvieron nos ayudan a evidenciar que los modelos basados en Machine Learning presentan una mayor capacidad para capturar patrones estacionales

complejos y no lineales superando a los enfoques estadísticos tradicionales en precisión y estabilidad. Toda esta implementación es clave para fortalecer la planificación energética del país, reducir problemas o riesgos de racionamiento y tomar mejores decisiones estratégicas en el sector eléctrico ecuatoriano.

*Palabras clave:* machine learning, series temporales, XGBoost, planificación energética, predicción de demanda energética, electric forecast, Z-Score, metodología KDD, MAE, RMSE, MAPE.



## ABSTRACT

This thesis comprises a comparative evaluation of forecasting algorithms applied to Ecuador's national energy demand, with the aim of identifying a model that enables us to estimate electricity consumption based on information collected between 2022 and 2025.

This research applies the KDD methodology, which follows a structured approach to pattern discovery in data by integrating the preparation, transformation, and modeling of time series. As part of the development, a prototype application named ELFO (Electric Forest) was proposed to automate the forecasting process, from ETL processing to the generation of demand predictions.

The data were obtained from official consumption reports, provided as Excel files with hourly records. These data require consolidation and cleaning to ensure consistency. During the transformation phase, anomaly detection procedures using Z-score were applied, along with missing-value handling, linear interpolation imputation, and finally daily temporal aggregation, which reduces high-frequency noise and facilitates the learning of seasonal trends.

For the modeling stage, a comparative approach was applied using two predictive architectures: an additive statistical model based on Facebook Prophet and a supervised Machine Learning model using XGBoost (with temporal feature engineering and hyperparameter optimization through Bayesian search).

Model performance was evaluated using standard metrics such as MAE, RMSE, and MAPE, complemented by comparative visualizations between actual and predicted values.

The results show that Machine Learning-based models have a greater capability to capture complex seasonal and nonlinear patterns, outperforming traditional statistical approaches in terms of accuracy and stability. This implementation is key to strengthening

national energy planning, reducing potential risk related to power rationing, and supporting better strategic decision-making in Ecuador's electricity sector.

*Keywords:* machine learning, series temporales, XGBoost, planificación energética, predicción de demanda energética, electric forecast, Z-Score, metodología KDD, MAE, RMSE, MAPE.

## **TABLA DE CONTENIDOS**

<b>CERTIFICACIÓN DE AUTORÍA.....</b>	<b>2</b>
<b>AUTORIZACIÓN DE DERECHOS DE PROPIEDAD INTELECTUAL .....</b>	<b>3</b>
<b>ACUERDO DE CONFIDENCIALIDAD .....</b>	<b>¡Error! Marcador no definido.</b>
<b>APROBACIÓN DE DIRECCIÓN Y COORDINACIÓN DEL PROGRAMA .....</b>	<b>4</b>
<b>DEDICATORIA .....</b>	<b>5</b>
<b>AGRADECIMIENTOS .....</b>	<b>7</b>
<b>RESUMEN.....</b>	<b>8</b>
<b>ABSTRACT.....</b>	<b>10</b>
<b>TABLA DE CONTENIDOS.....</b>	<b>12</b>
<b>ÍNDICE DE TABLAS.....</b>	<b>15</b>
<b>ÍNDICE DE FIGURAS.....</b>	<b>16</b>
<b>1. CAPÍTULO 1: INTRODUCCIÓN.....</b>	<b>18</b>
1.2    Justificación e importancia del trabajo de investigación.....	22
1.3    Alcance.....	26
1.4    Objetivos .....	29
1.4.2    Objetivos específicos .....	29
<b>2. CAPÍTULO 2: REVISIÓN DE LITERATURA .....</b>	<b>30</b>
2.1    Estado del arte .....	30
2.2    Marco teórico .....	38

2.2.1	Fundamentos de sistemas eléctricos y demanda energética .....	38
2.2.2	Series temporales y predicción energética .....	40
2.2.3	Algoritmos de Machine Learning para predicción energética .....	42
2.2.4	Herramientas para ciencia de datos.....	45
2.2.5	Metodologías de ciencia de datos .....	50
<b>3.</b>	<b>CAPÍTULO 3: DESARROLLO.....</b>	<b>56</b>
3.1	Metodología .....	56
3.1.1	Selección de Datos .....	59
3.1.2	Preprocesamiento de Datos.....	61
3.1.3	Transformación .....	62
3.1.4	Minería de datos.....	66
<b>4.</b>	<b>CAPÍTULO 4: ANÁLISIS DE RESULTADOS .....</b>	<b>71</b>
4.1	Pruebas de concepto .....	71
4.2	Ingesta de datos en ELFO .....	72
4.3	Entrenamiento .....	77
4.4	Evaluación.....	78
4.5	Proyección futura .....	79
4.6	Análisis de resultados.....	80
<b>5.</b>	<b>CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>84</b>
5.1	Conclusiones .....	84

5.2	Recomendaciones.....	85
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>		<b>88</b>
<b>ANEXOS.....</b>		<b>100</b>

## ÍNDICE DE TABLAS

Tabla 1 <i>Componentes fundamentales de una serie temporal</i> .....	41
Tabla 2 <i>Comparación general de algoritmos aplicados a predicción energética</i> .....	44
Tabla 3 <i>Comparación de metodologías de ciencia de datos</i> .....	55
Tabla 4 <i>Métricas de evaluación de modelos</i> .....	81

## ÍNDICE DE FIGURAS

Figura 1 <i>Curva de carga</i> .....	39
Figura 2 <i>Descripción general de los pasos que componen el proceso KDD</i> .....	52
Figura 3 <i>Modelo SEMMA</i> .....	53
Figura 4 <i>Etapas de metodología CRISP-DM</i> .....	54
Figura 5 <i>Arquitectura y flujo de información.</i> .....	58
Figura 6 <i>Muestra de reporte diarios de consumo</i> .....	59
Figura 7 <i>Tabla de reporte de producción de energía nacional.</i> .....	60
Figura 8 <i>Extracción y validación de fechas</i> .....	62
Figura 9 <i>Cálculo de Z-score</i> .....	64
Figura 10 <i>Manejo de valores nulos</i> .....	64
Figura 11 <i>Interpolación lineal</i> .....	65
Figura 12 <i>Resampling temporal</i> .....	66
Figura 13 <i>Configuración de feriados en modelo Prophet</i> .....	67
Figura 14 <i>Configuración de modelo XGBoost</i> .....	67
Figura 15 <i>Ingeniería de características</i> .....	68
Figura 16 <i>Optimización de búsqueda bayesiana</i> .....	69
Figura 17 <i>Carga de datos en ELFO</i> .....	73
Figura 18 <i>Procesamiento de archivos</i> .....	74
Figura 19 <i>Resumen de carga de datos</i> .....	75

Figura 20 <i>Almacenamiento de dataset en base de datos</i> .....	76
Figura 21 <i>Panel de configuración para entrenamiento</i> .....	77
Figura 22 <i>Panel de evaluación de resultados</i> .....	78
Figura 23 <i>Panel de evaluación de resultados</i> .....	79
Figura 24 <i>Proyección de consumo a 180 días</i> .....	80
Figura 25 <i>Modelo ganador</i> .....	82



## 1. CAPÍTULO 1: INTRODUCCIÓN

### 1.1 Definición del proyecto

El proyecto titulado “Evaluación y Comparativa de Algoritmos de Machine Learning para la Predicción de la Demanda Energética Nacional” se plantea como una iniciativa orientada a comprender y modelar el comportamiento histórico del consumo eléctrico del país. En un contexto donde la disponibilidad de energía es un recurso crítico para el funcionamiento de la industria, el comercio y el hogar, pronosticar adecuadamente la demanda se convierte en un mecanismo estratégico para la toma de decisiones gubernamentales y empresariales.

El proyecto pretende desarrollar un modelo predictivo de series temporales que permita estimar la demanda energética nacional con base en patrones históricos. Para ello, se consideran como insumos principales los datos registrados en los últimos tres años, periodo que se considera representativo para capturar tendencias estacionales, cambios estructurales y variaciones asociadas al comportamiento socioeconómico del país.

Desde una perspectiva científica y metodológica, el uso de algoritmos de Machine Learning para la predicción de la demanda energética responde a la necesidad de abordar fenómenos complejos caracterizados por relaciones no lineales, alta variabilidad temporal y múltiples factores de influencia simultánea. A diferencia de los métodos tradicionales de pronóstico, estos algoritmos permiten aprender directamente de los datos históricos, adaptarse a cambios estructurales en el comportamiento del consumo y mejorar progresivamente su capacidad predictiva a medida que se dispone de mayor información. La literatura especializada señala que, en contextos nacionales, la comparación entre distintos algoritmos no solo incrementa la precisión de las estimaciones, sino que también permite identificar modelos más robustos y generalizables frente a escenarios de incertidumbre energética. En este sentido, la evaluación comparativa propuesta en este proyecto se alinea con las

tendencias actuales de investigación en sistemas energéticos, aportando rigor científico y relevancia práctica al proceso de toma de decisiones en la planificación eléctrica (Hatie et al.,2017).

Se conoce que el consumo eléctrico por habitante en Ecuador experimentó un crecimiento sostenido entre 2009 y 2019, según los datos del Balance Energético Nacional de 2019. Durante esta década, el consumo per cápita aumentó de 1,088 kWh a 1,517 kWh, lo que representa un incremento total del 39.4%. Además, el Instituto de Investigación Geológico y Energético detalló que este crecimiento continuó en el corto plazo, registrándose un aumento del 2% entre 2018 y 2019, al pasar el consumo de 1,488 kWh a los 1,517 kWh finales por habitante (IIGE, 2020).

La capacidad de anticipar la demanda energética con un alto nivel de precisión permite optimizar la planificación del sistema eléctrico, ajustando la generación de electricidad a las necesidades reales. Esto implica una mejor gestión del despacho energético, reduciendo pérdidas, evitando sobrecostos operativos y fortaleciendo la seguridad energética nacional en escenarios de alta variabilidad.

Asimismo, el presente proyecto se define como una investigación aplicada que busca integrar el análisis de datos históricos con técnicas avanzadas de aprendizaje automático para generar conocimiento útil y accionable en el ámbito energético. Su enfoque comparativo permite no solo medir el desempeño de distintos algoritmos, sino también establecer criterios técnicos que orienten su posible implementación en contextos reales. De este modo, el estudio se concibe como un aporte que articula la dimensión académica con las necesidades operativas del sistema eléctrico nacional, contribuyendo al fortalecimiento de la planificación energética y al desarrollo de estrategias orientadas a la eficiencia, sostenibilidad y seguridad del suministro eléctrico a mediano y largo plazo (International Energy Agency,2022).

Uno de los componentes esenciales del proyecto es la elección de algoritmos de Machine Learning, lo que permitirá determinar cuál técnica ofrece un mejor rendimiento en términos de precisión, estabilidad y capacidad de generalización. Entre los algoritmos candidatos podrían considerarse modelos como LSTM, CNN o MLP, cada uno con características particulares para capturar patrones temporales complejos.

La comparación rigurosa entre estos modelos permitirá no solo identificar cuál técnica se adapta mejor a la dinámica de la demanda energética nacional, sino también proponer mejoras y ajustes que puedan integrarse a futuros sistemas de gestión energética. El análisis incluirá métricas de error como MAE, RMSE y MAPE, fundamentales para evaluar el desempeño de predicción.

Además del componente técnico, el proyecto se orienta a generar un insumo estratégico para las instituciones responsables de la planificación eléctrica, incluyendo ministerios, operadores de red y empresas distribuidoras. Los resultados obtenidos permitirán diseñar políticas más eficientes, estimar necesidades de ampliación de infraestructura y anticipar escenarios de tensión energética.

Un aspecto clave dentro de la definición del proyecto es su contribución a la gestión de racionamientos eléctricos, especialmente en épocas de estiaje, donde la disponibilidad de recursos hídricos disminuye significativamente. Al contar con un modelo predictivo robusto, las autoridades podrán planificar medidas preventivas con mayor anticipación, reduciendo los impactos sociales y económicos asociados a la escasez energética.

El análisis de series temporales ofrece una ventana al comportamiento cíclico de la demanda, permitiendo identificar patrones diarios, semanales y estacionales. El proyecto busca capturar estas dinámicas con el fin de ofrecer un pronóstico que no solo sea numéricamente preciso, sino también contextualmente interpretable, facilitando la toma de decisiones por parte de actores técnicos y administrativos.

Desde el punto de vista científico, este enfoque permite abordar el fenómeno de la demanda energética como un sistema dinámico complejo, en el que interactúan factores técnicos, económicos, sociales y ambientales. La combinación de análisis de series temporales con algoritmos de Machine Learning se sustenta en paradigmas actuales de la ciencia de datos, que priorizan la extracción de conocimiento a partir de datos reales y la validación empírica de los modelos predictivos. Este enfoque contribuye a reducir la brecha entre la modelación teórica y la aplicación práctica, fortaleciendo la validez externa de los resultados y su potencial de transferencia a contextos institucionales reales. En este sentido, la investigación se alinea con las tendencias contemporáneas en estudios energéticos, que promueven soluciones basadas en datos como soporte fundamental para la planificación y gestión de sistemas eléctricos nacionales (Hastie et al., 2017).

La implementación del modelo requerirá un tratamiento adecuado de los datos históricos, considerando procesos como limpieza, normalización, imputación de valores faltantes y detección de anomalías. Estos pasos son esenciales para garantizar la calidad del dataset y asegurar que los algoritmos analicen información confiable y consistente.

Otro componente relevante del proyecto es la visualización de resultados, que permitirá evaluar de manera intuitiva el comportamiento histórico y proyectado de la demanda energética. Gráficos comparativos entre predicciones y valores reales facilitarán la validación de los modelos y servirán como una herramienta de comunicación para los diferentes grupos de interés.

Este proyecto también busca sentar las bases para la creación de una plataforma de predicción continua, donde el modelo pueda actualizarse de forma periódica con nuevos datos y mantener su capacidad predictiva a lo largo del tiempo. Este enfoque contribuiría al establecimiento de un sistema de monitoreo dinámico del consumo energético nacional.

Finalmente, la definición del proyecto destaca su importancia como contribución tecnológica y estratégica para el país. El uso de técnicas modernas de Machine Learning aplicadas al sector energético permitirá avanzar hacia una gestión más eficiente, sostenible y resiliente del sistema eléctrico, beneficiando tanto a las instituciones responsables como a la ciudadanía que depende del suministro constante de energía.

## **1.2 Justificación e importancia del trabajo de investigación**

La predicción precisa de la demanda energética nacional constituye un insumo estratégico para la seguridad energética, la planificación del despacho y la optimización de recursos. Sin pronósticos fiables, los operadores del sistema y los planificadores enfrentan mayores riesgos de desbalance entre generación y consumo, con consecuencias económicas y sociales significativas (IEA, 2025).

A medida que la electrificación de sectores como transporte y servicios se acelera, la demanda eléctrica global ha mostrado crecimientos sostenidos en los últimos años, lo que obliga a mejorar las herramientas de pronóstico para evitar déficits o sobrecapacidad innecesaria (Ember, 2025). Incorporar modelos predictivos robustos permite al país anticipar picos y planificar inversiones en infraestructura con mayor eficiencia.

La justificación técnica de este trabajo radica en la evidencia creciente de que técnicas modernas de Machine Learning como LSTM, CNN, MLP y modelos híbridos que superan en muchos casos a los métodos clásicos en precisión y capacidad de captura de patrones no lineales en series temporales de carga (Choi et al., 2020, p. 1109). Evaluar y comparar estas técnicas en el contexto nacional aportará conocimiento práctico sobre su aplicabilidad real.

Desde el punto de vista operativo, mejores pronósticos de demanda permiten optimizar el despacho y las reservas del sistema, reduciendo costos operacionales y las emisiones asociadas a la generación de respaldo fósil (Fose et al., 2024). Esto es

particularmente relevante cuando la matriz incluye una alta proporción de renovables intermitentes.

La investigación tiene una fuerte componente socioeconómica: pronósticos más precisos contribuyen a minimizar cortes y racionamientos, que impactan negativamente la productividad industrial, la educación y la calidad de vida. En regiones con vulnerabilidad hídrica o dependencia de hidroelectricidad, la capacidad predictiva puede ser la diferencia entre una gestión ordenada y racionamientos abruptos (Adu-Poku et al., 2024).

En esta perspectiva, la mejora en la precisión de los pronósticos de demanda energética tiene implicaciones directas sobre la equidad, el bienestar y la estabilidad social. La reducción de escenarios de racionamiento o planificación deficiente del suministro energético contribuye a minimizar impactos negativos sobre los hogares, los servicios esenciales y los sectores productivos más vulnerables. En este sentido, la investigación trasciende el ámbito técnico para situarse como un aporte al desarrollo sostenible, al fortalecer la capacidad del Estado y de los operadores energéticos para anticipar riesgos y diseñar políticas preventivas basadas en evidencia. Diversos estudios señalan que la planificación energética informada es un factor clave para reducir desigualdades y mejorar la resiliencia de los sistemas socioeconómicos frente a escenarios de incertidumbre (CEPAL, 2020).

A nivel de política pública, los resultados del estudio ofrecerán evidencia cuantitativa para diseñar medidas preventivas en periodos de tensión (por ejemplo, incentivos para ahorro, esquemas de demanda flexible o acuerdos comerciales internacionales de importación de energía) (Pérez-López et al., 2025, p. 5217). Un modelo validado por datos facilita la toma de decisiones basada en evidencia.

La justificación académica está apoyada por la brecha identificada en la literatura: aunque existen reseñas y estudios aplicados, son necesarias comparativas locales que

consideren la idiosincrasia de los datos nacionales (estacionalidad, rupturas por políticas, efectos socioeconómicos) para seleccionar modelos realmente útiles en la práctica (Hasan et al., 2025).

Además, este trabajo contribuirá a la transferencia tecnológica: documentar pipelines de preprocesamiento, selección de variables y arquitectura de modelos facilita la implementación operativa por parte de operadores y centros de investigación nacionales. Esa transferencia acelera la adopción de soluciones data-driven en el sector energético (Chen et al., 2025).

La investigación también está justificada por razones climáticas: cambios en patrones de temperatura y precipitación afectan tanto la demanda como la oferta (capacidad de hidroelectricidad), por lo que modelos que integren información climática mejoran la resiliencia del sistema (IPCC, 2023). Incorporar variables climáticas y escenarios de cambio climático incrementa la relevancia de las predicciones.

Desde una perspectiva económica, optimizar la programación de generación y reducir la incertidumbre de demanda disminuye los costos marginales de producción, evita cargas regulatorias y puede reducir el precio final para consumidores y la exposición al mercado spot. El impacto económico positivo justifica la inversión en investigación aplicada.

El estudio aportará metodologías contrastables: al comparar algoritmos con métricas estandarizadas (MAE, RMSE, MAPE) y pruebas de robustez (validación cruzada) se creará un marco reproducible que otros investigadores y técnicos podrán replicar. Esto mejora la trazabilidad científica de las decisiones operativas.

La inclusión de técnicas de Machine Learning permite capturar efectos no lineales y dependencias temporales largas (por ejemplo, LSTM), lo cual es crucial para series con memoria y patrones compuestos; por tanto, su evaluación es pertinente para maximizar la precisión en horizontes de corto y mediano plazo (Huang & Yi, 2024).

A nivel regulatorio, esta investigación puede apoyar la definición de requisitos mínimos para herramientas de forecasting que los agentes del mercado deben proveer, contribuyendo a una gobernanza más robusta del sector eléctrico y a una mejor coordinación entre actores. La estandarización favorece la competencia y la transparencia.

La justificación metodológica incluye la necesidad de trabajar con datos de tres años, como propone el proyecto, puesto que este horizonte captura múltiples ciclos estacionales y eventos atípicos que permiten entrenar modelos con mayor capacidad de generalización frente a variaciones anuales. Sin embargo, la metodología considerará también estrategias para extender o limitar ventanas temporales según resultados empíricos (Hasan et al., 2025).

En términos de sostenibilidad, mejores pronósticos favorecen la integración de renovables al permitir una programación más eficiente de fuentes térmicas y almacenamiento, reduciendo el uso de combustibles fósiles como respaldo y ayudando a cumplir metas de descarbonización (Ember, 2025). Este beneficio ambiental aporta un argumento adicional para la investigación.

La investigación tiene impacto en la gestión de riesgos: un modelo que incluya intervalos de confianza y estimaciones probabilísticas permitirá a operadores y reguladores prepararse ante escenarios extremos (picos de demanda o fallas en la generación) (Brailey, 2024). La predicción probabilística mejora la toma de decisiones bajo incertidumbre.

La dimensión social de la investigación no es menor: evitar racionamientos o planificarlos con antelación reduce la vulnerabilidad de hogares y sectores críticos (salud, agua, telecomunicaciones) y protege a los estratos más afectados por cortes prolongados (Adu-Poku et al., 2024). Por ello, la mejora en forecasting tiene un claro componente de bienestar público.

La robustez del estudio se garantizará mediante validaciones fuera de muestra y pruebas en diferentes horizontes (horas, días, semanas), lo que permitirá recomendar



soluciones operativas específicas para distintos plazos temporales de toma de decisiones (Song et al., 2025). Esta versatilidad incrementa la utilidad del proyecto para operadores y planificadores.

Adicionalmente, la comparación entre modelos favorecerá la identificación de enfoques híbridos (por ejemplo, combinaciones de ML con métodos estadísticos) que puedan equilibrar interpretabilidad y precisión, una necesidad recurrente en contextos donde las decisiones deben ser auditables.

Desde la óptica del mercado eléctrico, mejores predicciones reducen la volatilidad de precios y limitan la exposición de agentes a mercados spot, contribuyendo a la estabilidad macroeconómica y financiera del sector. Esta externalidad económica amplía la relevancia del proyecto.

La investigación también está justificada por recientes eventos internacionales que muestran cómo condiciones meteorológicas adversas (sequías, olas de calor) pueden forzar racionamientos o aumentos de importaciones de energía, subrayando la necesidad de modelos predictivos integrados con variables climáticas y de disponibilidad de recursos. Estas lecciones internacionales refuerzan la pertinencia del estudio.

En suma, la justificación del proyecto combina argumentos técnicos, económicos, sociales y ambientales: un pronóstico de demanda más preciso mejora la operación del sistema, protege a la población frente a racionamientos, facilita la integración de renovables y aporta evidencia para políticas públicas eficaces; por ello, la investigación propuesta es prioritaria y de alto valor estratégico.

### **1.3 Alcance**

El alcance de este proyecto comprende el desarrollo y evaluación de algoritmos de Machine Learning aplicados a la predicción de la demanda energética nacional. Esto implica analizar el comportamiento histórico del consumo eléctrico en un periodo de tres años y

generar modelos capaces de anticipar la demanda en horizontes de corto y mediano plazo. La investigación se centra únicamente en el análisis de datos de demanda, sin intervenir en la infraestructura física del sistema eléctrico.

El estudio abarcará la recopilación, depuración y procesamiento de datos provenientes de fuentes oficiales, garantizando la calidad y consistencia del conjunto de datos. Esto incluye el tratamiento de valores faltantes, la detección de anomalías y la normalización de series para asegurar un entrenamiento adecuado de los modelos. El alcance contempla solo datos de demanda energética que influyan directamente en el consumo.

Dentro del alcance técnico se incluye la implementación de algoritmos de predicción de series temporales como métodos basados en aprendizaje automático (Random Forest) y modelos avanzados basados en redes neuronales (LSTM). La selección final de algoritmos se basará en su relevancia para escenarios energéticos y su capacidad comprobada para capturar patrones complejos.

El proyecto contempla el diseño de un pipeline estandarizado de modelado, que abarque desde la preparación de datos hasta la validación y comparación de resultados. Esto incluye el uso de técnicas como división temporal del conjunto de datos y validación cruzada basada en tiempo. Para finalizar el alcance incluye la creación de una plataforma web local interactiva que permita gestionar los ejercicios de predicción que se quiera implementar bajo demanda del o los usuarios de este proyecto de investigación, los resultados podrán ser utilizados como base para futuras implementaciones.

Se evaluará el rendimiento de los modelos utilizando métricas cuantitativas ampliamente aceptadas en el ámbito del pronóstico de demanda eléctrica, tales como MAE, RMSE, MAPE. Asimismo, se generarán visualizaciones comparativas que permitan interpretar el comportamiento de cada modelo respecto a los datos reales. El alcance se limita

a la evaluación académica y técnica, sin incluir decisiones operativas específicas para el sistema eléctrico.

El trabajo comprenderá también un análisis comparativo que permita identificar el modelo más adecuado según los criterios de precisión, estabilidad, capacidad de generalización y facilidad de implementación. Este análisis proporcionará recomendaciones que puedan servir de guía para operadores eléctricos, instituciones públicas o investigadores interesados en modelos predictivos de demanda.

La investigación se limitará a la demanda energética nacional en su conjunto, sin realizar desagregaciones por regiones, sectores económicos o niveles de voltaje. Si bien dichos análisis podrían aportar información más granular, se encuentran fuera del alcance debido a restricciones de datos y a la naturaleza global del estudio.

El proyecto no abarca simulaciones del sistema eléctrico, modelado de generación, análisis de costos marginales ni estudios de mercado eléctrico. Su alcance se restringe estrictamente a la predicción del comportamiento de la demanda y a la comparación de algoritmos que permitan realizar dicho pronóstico con eficiencia.

El alcance del proyecto se limita al análisis y predicción de la demanda eléctrica a partir de datos históricos consolidados, sin considerar variables externas relacionadas con la infraestructura física del sistema eléctrico, como capacidad de generación, transmisión o distribución. En consecuencia, los resultados obtenidos deben interpretarse como un apoyo a la planificación y análisis de la demanda, mas no como una herramienta de simulación integral del sistema eléctrico nacional.

Asimismo, el estudio no contempla la implementación de los modelos en un entorno productivo ni su integración directa con sistemas operativos de gestión energética. Las predicciones generadas se enfocan exclusivamente en el análisis académico y experimental,

por lo que no se evalúan aspectos como la latencia, disponibilidad o escalabilidad del sistema en escenarios reales de operación continua.

Finalmente, el alcance del proyecto se restringe al análisis del consumo eléctrico a nivel agregado, sin desagregar por sectores específicos como residencial, industrial o comercial. Si bien este enfoque permite una visión global del comportamiento de la demanda, se reconoce que estudios futuros podrían beneficiarse de un análisis segmentado que permita capturar patrones particulares asociados a distintos tipos de usuarios. Estas limitaciones no afectan la validez del estudio, sino que delimitan claramente el contexto en el cual deben interpretarse los resultados obtenidos.

## **1.4 Objetivos**

### ***1.4.1 Objetivo general***

Determinar un modelo de aprendizaje automático que ayude a predecir de mejor manera el consumo energético del Ecuador mediante el análisis de variables históricas de demanda energética nacional desde 2022 hasta 2025, con el fin de gestionar y maximizar la disponibilidad de electricidad para la población ecuatoriana.

### ***1.4.2 Objetivos específicos***

Recolectar la data de demanda energética del Ecuador desde enero 2022 hasta septiembre de 2025.

Realizar el análisis exploratorio de datos EDA en el conjunto de datos de consumo energético nacional.

Analizar el rendimiento de distintos modelos de aprendizaje automático para la predicción de consumo energético nacional.

Evaluar el rendimiento de los modelos implementados para la predicción de la demanda energética nacional.

## 2. CAPÍTULO 2: REVISIÓN DE LITERATURA

### 2.1 Estado del arte

El sistema eléctrico del Ecuador ha sido históricamente influenciado por factores climáticos, hidrológicos y socioeconómicos. De acuerdo con la Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables (ARCERNNR, 2022), más del 80% de la generación eléctrica nacional proviene de fuentes hídricas, lo que convierte al sistema en altamente sensible a eventos como estiajes, sequías o retrasos en temporadas de lluvias.

El interés por modelar y predecir la demanda eléctrica en Ecuador tiene raíces administrativas y técnicas; los informes oficiales de balance y planificación energética del país describen la necesidad de mejorar herramientas de proyección para asegurar el abastecimiento, identificar inversiones y diseñar políticas de contingencia. Documentos institucionales como el Balance Energético Nacional de la Corporación Eléctrica Nacional (CELEC) y reportes del Operador Nacional de Electricidad (CENACE) y de la Agencia de Regulación y Control de Electricidad (ARCONEL) ofrecen las bases de datos y las metodologías de proyección utilizadas por los planificadores nacionales (CENACE, 2025). Estas fuentes no solo contienen estadísticas históricas, sino también criterios metodológicos que han guiado estudios académicos posteriores.

En el ámbito académico, varios trabajos de pregrado y posgrado en universidades ecuatorianas han abordado el pronóstico de la demanda eléctrica usando técnicas estadísticas y de Machine Learning. Por ejemplo, tesis y artículos de la Universidad Politécnica Salesiana (UPS) y la Universidad de las Fuerzas Armadas (ESPE) presentan modelos SARIMA (Jaramillo, 2024), regresiones multivariantes (Ortiz & García, 2015) y acercamientos de minería de datos (Jaramillo & Llamuca, 2022) aplicados a series de consumo en diferentes escalas territoriales. Estas investigaciones prueban la aplicabilidad de métodos clásicos y adaptados al contexto local.

Mora y Toaquiza (2021) también aplica modelos ARIMA y de redes neuronales para pronosticar la demanda eléctrica ecuatoriana, concluyendo que las redes neuronales ofrecían una mayor precisión en períodos de alta variabilidad. Por su parte, Sánchez y Sinche, (2024) emplearon modelos machine learning para optimizar predicciones en regiones específicas, destacando la importancia de integrar variables climáticas.

Chicaiza y Carrillo (2025) propusieron modelos de Long Short Term Memory (LSTM) para pronosticar la demanda horaria de Ecuador, logrando disminuir significativamente el error respecto a métodos estadísticos tradicionales. Sus resultados evidencian la utilidad de técnicas profundas en un sistema con alta dependencia climática.

A pesar de la diversidad de enfoques identificados en la literatura, se evidencia que una parte significativa de los estudios, tanto internacionales como nacionales, se centra en la aplicación de modelos individuales o en evaluaciones parciales de desempeño, sin desarrollar comparaciones sistemáticas entre distintos algoritmos bajo un mismo marco de datos y validación. Asimismo, varios trabajos priorizan horizontes temporales específicos (horario o diario), lo que limita la comprensión integral del comportamiento de la demanda a escala nacional. Esta situación revela un vacío metodológico en relación con la evaluación comparativa de algoritmos de Machine Learning aplicados a series temporales energéticas de largo plazo, particularmente en contextos con alta dependencia hidroeléctrica y variabilidad climática como el ecuatoriano. La identificación de este vacío justifica la pertinencia del presente estudio y delimita claramente su aporte dentro del campo de investigación energética (Hong et al. 2016; Shahid et al., 2020).

Un aporte notable en la literatura nacional es el estudio sobre data mining y pronóstico a corto plazo publicado en la revista técnica de CENACE, el cual propone herramientas computacionales en Python para minería de datos y forecasting, enfatizando la importancia de variables exógenas (temperatura, días festivos, actividad económica) y

validación temporal para series nacionales (Gallo et al., 2021). Este trabajo sirvió para demostrar que, con ajustes adecuados, modelos de aprendizaje automático pueden superar a métodos puramente estadísticos en ciertos horizontes de predicción.

CENACE (2018) también desarrolló modelos de proyección anual basados en tendencias y crecimiento económico, sin embargo, dichos enfoques presentan limitaciones para capturar dinámicas diarias o intradiarias.

Investigaciones a nivel local muestran que, en los sistemas de distribución, a menor escala, también se han obtenido resultados prometedores. Un estudio sobre la Empresa Eléctrica de Ambato (EEASA) documentó la implementación de criterios de Big Data para predicción en la red de distribución, evidenciando los retos de calidad de datos y la necesidad de agrupamiento de usuarios para mejorar la predicción a nivel de alimentadores y subestaciones (Guaman et al., 2019). Este tipo de estudios locales es crucial para trasladar modelos nacionales a operaciones distribuidas.

En el plano metodológico, trabajos recientes provenientes de Ecuador exploran la combinación de técnicas tradicionales y enfoques de *Machine Learning*. Por ejemplo, comparativas entre modelos SARIMA y regresiones con variables exógenas han sido implementadas para el Sistema Nacional Interconectado (SNI), permitiendo identificar estacionalidades fuertes y rupturas en la serie (Suntaxi et al., 2019). Estas evaluaciones metodológicas han sido publicadas en revistas nacionales y repositorios académicos, aportando marcos de validación reproducibles.

La pandemia de COVID-19 generó rupturas en los patrones de consumo eléctrico que fueron objeto de análisis por investigadores ecuatorianos; estudios que abordaron los primeros dos años de la pandemia en Quito mostraron la necesidad de modelos adaptativos capaces de incorporar cambios estructurales repentinos y de reestimación dinámica de parámetros, algo que modelos estáticos no capturan adecuadamente (Arévalo, 2024). Estas

lecciones han impulsado el interés por modelos con actualización continua y por metodologías de detección de cambios.

Si bien los estudios a escala local y sectorial aportan información valiosa para la gestión descentralizada del consumo eléctrico, la literatura revisada muestra una limitada integración de estos enfoques en modelos de alcance nacional que permitan apoyar la planificación estratégica del sistema eléctrico en su conjunto. La extrapolación de resultados locales hacia decisiones de política energética requiere marcos metodológicos robustos y comparables, capaces de capturar tanto patrones agregados como variaciones estructurales en el tiempo. En este contexto, la revisión de la literatura pone de manifiesto la necesidad de investigaciones que articulen el análisis de datos históricos nacionales con técnicas avanzadas de aprendizaje automático, garantizando resultados transferibles y útiles para la toma de decisiones a nivel institucional y regulatorio (International Energy Agency, 2022).

Algunos trabajos han aplicado redes neuronales y metaheurísticas para optimizar la predicción en escenarios locales. Por ejemplo, investigaciones que usan optimización por *Grey Wolf Optimization* (GWO) para ajustar regresores o redes han mostrado mejoras en el ajuste de modelos para series locales cuando los datos son ruidosos o tienen registros faltantes (Martins & Macêdo, 2024). Estos estudios demuestran que la hibridación (algoritmos de optimización + Machine Learning) es una vía fructífera para situaciones con datos limitados o de mala calidad.

A nivel provincial y cantonal existen aplicaciones prácticas del *forecasting* eléctrico; trabajos sobre consumo por cantones en Pichincha documentan aproximaciones que combinan variables demográficas y actividad local con redes neuronales simples, alcanzando errores relativos aceptables para planificación local (Guachimboza-Davalos et al., 2021). Estas investigaciones resaltan que la granularidad geográfica mejora la comprensión de patrones de demanda y puede informar políticas municipales de eficiencia energética.



El uso de series temporales y técnicas de *clustering* de consumidores ha sido explorado para mejorar la predicción a nivel de usuario y facilitar la creación de perfiles de consumo. Estudios de tesis en universidades ecuatorianas han propuesto agrupar patrones de uso residencial y comercial para luego aplicar modelos por cluster, lo que reduce la varianza de las predicciones y permite estrategias de demanda flexible más dirigidas (Chong & Aguilar, 2016). Estas experiencias locales son útiles para servicios públicos que buscan programas de gestión de la demanda.

Los informes sectoriales y regulatorios (ARCONEL) han documentado cómo eventos climáticos y políticas energéticas han alterado los balances de energía, aportando datos empíricos que han sido usados por investigadores para validar modelos de predicción que integran variables climáticas y operativas (ARCONEL, 2025). El contraste entre proyecciones regulatorias y resultados empíricos ha incentivado la investigación aplicada para reducir desviaciones en estimaciones de mediano plazo.

Estudios sobre impacto hidro climático en la generación eléctrica (relevantes para Ecuador por su fuerte dependencia hídrica) han impulsado la inclusión de índices hidrológicos y climatológicos como entradas en modelos predictivos de demanda y oferta. La literatura nacional, complementada con análisis internacionales, sugiere que incorporar variables de disponibilidad hídrica mejora la gestión ante sequías y puede reducir la incertidumbre en escenarios críticos (Stott, 2024). Esto es especialmente relevante luego de episodios recientes de racionamiento.

Investigaciones recientes se han enfocado en la predicción de demanda en contextos de estiaje, especialmente dado el impacto de fenómenos como El Niño y La Niña. Chamorro y Mera (2025) analizaron cómo las variaciones en el caudal de los ríos afectaban la disponibilidad de generación hidráulica, mostrando cómo la reducción de reservas obligaba a un uso más eficiente de los pronósticos de demanda.

Investigaciones publicadas en revistas universitarias ecuatorianas han propuesto metodologías de validación robustas (por ejemplo, validación cruzada temporal, intervalos de predicción) para asegurar que los modelos propuestos no solo ajusten bien en muestra, sino que generalicen fuera de muestra (Serrano-Guerrero et al., 2021). Estos esfuerzos metodológicos son centrales para trasladar resultados académicos a aplicaciones operativas confiables.

La disponibilidad y calidad de datos es un desafío recurrente en la bibliografía ecuatoriana: múltiples trabajos señalan la necesidad de mejorar telemetría, resolución temporal y registros meteorológicos correlacionados pero centralizados. Sin estos insumos, los modelos avanzados tienden a sobre ajustar o a perder capacidad predictiva fuera de los rangos observados, por lo que muchos estudios proponen protocolos de limpieza e imputación específicos para los datos nacionales (Candelario & Albán, 2024).

Además de modelos puntuales, la literatura ecuatoriana ha empezado a explorar predicciones probabilísticas e intervalos de confianza para la demanda, buscando entregar no solo un valor esperado sino estimaciones de incertidumbre que ayuden a la toma de decisiones operativas (reservas, despacho, contratos de compra). Investigaciones asociadas con metodologías de intervalos de predicción han sido adaptadas en contextos locales y demuestran utilidad práctica en la planificación de contingencias (Serrano-Guerrero et al., 2021).

Diversos trabajos de la última década han evidenciado la conveniencia de modelos híbridos que combinan razonamiento físico (por ejemplo, efectos conocidos de temperatura sobre consumo) con aprendizaje automático para captar patrones residuales complejos. En Ecuador, propuestas de este tipo han mostrado mejores prestaciones en horizontes de corto plazo que modelos puramente empíricos o puramente físicos. Esto sugiere que la integración de conocimiento del dominio local es ventajosa (Radicelli & Parra, 2025).

Los artículos y tesis que analizan la demanda por sectores (residencial, industrial, transporte) resaltan la heterogeneidad del comportamiento de consumo y la necesidad de modelos desagregados para políticas sectoriales. En Ecuador, la composición del consumo por sectores ha cambiado en años recientes, lo que exige modelos que puedan incorporar cambios estructurales en la economía y en la adopción de tecnologías (CELEC, 2022).

El impacto de eventos extremos y fallos de infraestructura (por ejemplo, el apagón nacional de 2024 y la intensificación de racionamientos por sequía) ha subrayado la urgencia de sistemas predictivos con alertas tempranas (AP, 2024). La literatura y noticias nacionales ponen en evidencia cómo la falta de modelos integrados de demanda-oferta puede amplificar impactos sociales y económicos, motivando líneas de investigación orientadas a la resiliencia.

En el ámbito de transferencia tecnológica, algunos trabajos ecuatorianos han desarrollado prototipos y herramientas en Python y plataformas abiertas para que operadores locales puedan reproducir análisis de forecasting; esta labor conjunta entre academia y operadores fortalece la implementación práctica de modelos y facilita la capacitación de personal técnico (Montero Laurencio et al., 2024). Publicaciones de repositorios universitarios y talleres técnicos documentan estas iniciativas.

Los estudios comparativos publicados en Ecuador también abordan el horizonte temporal óptimo para diferentes aplicaciones: mientras que para despacho intradiario son preferibles modelos con alta resolución, para planificación de mediano plazo modelos mensuales o anuales robustos y menos sensibles a ruido resultan más adecuados (Salazar & Panchi, 2014). La literatura local aporta recomendaciones prácticas sobre ventanas de entrenamiento y frecuencia de actualización de modelos.

La revisión muestra, además, iniciativas recientes que incorporan variables económicas y de movilidad para capturar cambios en la demanda asociada a la recuperación postpandemia y a cambios en patrones de trabajo. Estos enfoques multivariantes han

mostrado mejoras en la precisión cuando se disponen de series adicionales (actividad económica o indicadores comerciales) que correlacionan con el consumo eléctrico. (Estrella & Pazuña, 2021)

Un aspecto emergente en la literatura ecuatoriana es la exploración de forecasting multi-energía y la integración de la demanda eléctrica con otras demandas energéticas (combustibles, transporte), buscando una visión sistémica que permita políticas energéticas coherentes. Algunos trabajos usan herramientas prospectivas como LEAP para escenarios de largo plazo, complementando los modelos estadísticos de corto plazo (Tigua et al., 2025).

Igualmente, se evidencia avances significativos en el uso de técnicas estadísticas y de Machine Learning para la predicción de la demanda eléctrica; sin embargo, también pone de manifiesto limitaciones asociadas a la falta de comparaciones integrales entre modelos, a la escasa consideración de horizontes temporales amplios y a la necesidad de enfoques adaptados a contextos nacionales específicos. Estas brechas identificadas sustentan la relevancia del presente estudio, el cual se orienta a evaluar de manera comparativa distintos algoritmos de Machine Learning aplicados a la predicción de la demanda energética nacional, bajo un enfoque metodológico riguroso y reproducible, alineado con las necesidades actuales del sistema eléctrico ecuatoriano.

Finalmente, el estado del arte en Ecuador indica que, aunque existe un cuerpo creciente de trabajos locales sobre predicción de demanda eléctrica, hay áreas que requieren mayor investigación: validación cruzada entre instituciones, estandarización de conjunto de datos, desarrollo de predicción probabilística operativa y mayor colaboración entre regulador, operador y academia. Estas brechas son precisamente las que justifica y enmarca el proyecto de evaluación comparativa de algoritmos a escala nacional.

## 2.2 Marco teórico

### 2.2.1 *Fundamentos de sistemas eléctricos y demanda energética*

Los sistemas eléctricos están compuestos por subsistemas de generación, transmisión, distribución y consumo. Kundur (1994) indica que la estabilidad y eficiencia del sistema dependen de una coordinación precisa entre estos elementos, siendo el pronóstico de demanda un componente esencial para garantizar el equilibrio entre la oferta y el consumo energético.

La demanda eléctrica puede clasificarse en demanda base, demanda media y demanda pico. Este comportamiento depende de factores como el clima, la actividad económica, los hábitos domésticos y la disponibilidad de tecnologías de climatización (Gellings & Chamberlin, 1988). Comprender estas categorías permite estructurar estrategias de predicción que reflejen la dinámica real del comportamiento energético.

Los sistemas de generación deben planificar su operación mediante mecanismos de despacho económico y control de carga. Grainger y Stevenson (1994) explican que el despacho óptimo busca minimizar el costo de producción cumpliendo restricciones técnicas como límites de potencia, rampas máximas, y estabilidad en la red. Un pronóstico preciso ayuda a evitar estados de sobrecarga o déficit.

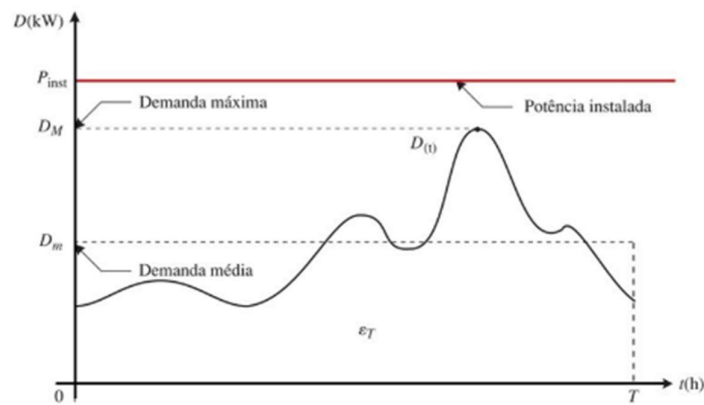
Los fenómenos climáticos también juegan un rol crítico en la demanda eléctrica. Silva et al. (2020) señalan que la temperatura, humedad y radiación solar pueden modificar el consumo residencial y comercial, especialmente en países con alta dependencia de climatización artificial.

El concepto de elasticidad energética resulta relevante para comprender la relación entre actividad económica y consumo eléctrico. Según Berndt (1991), al incrementarse la producción industrial o el poder adquisitivo, la demanda de electricidad también crece, aunque no necesariamente de manera proporcional.

Otro concepto fundamental es la curva de carga como se puede observar en la Figura 1, la cual representa la evolución de la demanda en un periodo determinado. Esta curva permite identificar picos, valles y períodos de estabilidad, siendo una de las herramientas más usadas para el diseño de modelos de predicción energética y análisis operativo del sistema.

**Figura 1**

*Curva de carga*



*Nota.* Elaboración propia.

Finalmente, el análisis de la demanda se complementa con indicadores como el factor de carga, el factor de potencia y el consumo per cápita, los cuales permiten evaluar la eficiencia del sistema y determinar necesidades futuras de infraestructura (Pansini, 2020).

Desde una perspectiva teórica, la representación de la demanda eléctrica mediante series temporales permite integrar los principios operativos del sistema eléctrico con el análisis cuantitativo del comportamiento del consumo a lo largo del tiempo. La curva de carga, la elasticidad de la demanda y los factores climáticos descritos previamente encuentran en las series temporales un marco formal que facilita su modelación, interpretación y predicción. En este sentido, el análisis temporal no se limita a una herramienta estadística, sino que se constituye como un puente conceptual entre el funcionamiento físico del sistema eléctrico y la toma de decisiones operativas y estratégicas. Esta articulación resulta

fundamental para comprender cómo las variaciones estructurales y estacionales del consumo impactan la planificación del despacho, la confiabilidad del sistema y la eficiencia económica del sector energético (Box et al,2016).

### **2.2.2 Series temporales y predicción energética**

Las series temporales constituyen una herramienta fundamental para el estudio de fenómenos que evolucionan de manera secuencial en el tiempo. Box et al. (2016) formalizaron un marco analítico que incluye los modelos AR, MA y ARIMA, proporcionando una base para analizar tendencias, estacionalidad y componentes aleatorios. Estos modelos han sido ampliamente utilizados en contextos energéticos debido a la naturaleza cíclica de la demanda eléctrica, la cual suele presentar picos diarios, semanales y estacionales (Hyndman y Athanasopoulos, 2018).

La predicción de la demanda energética se nutre de la capacidad de identificar patrones regulares y comportamientos recurrentes en los datos históricos. Según Richardson et al. (2010), la demanda eléctrica nacional refleja patrones dependientes del comportamiento humano, la variación climática, la productividad industrial y factores socioeconómicos. De esta forma, los métodos de series temporales buscan modelar no solo la dependencia temporal, sino también las relaciones multivariadas con factores externos.

La descomposición de series es un proceso esencial que permite separar una señal temporal en componentes estructurales: tendencia, estacionalidad y residuo. Winters (1960) desarrolló métodos de suavizamiento exponencial que facilitan esta descomposición, especialmente en sistemas donde la estacionalidad es pronunciada, como es usual en demandas eléctricas asociadas a climas tropicales y subtropicales. En tales sistemas, los patrones pueden ser influenciados por la luz solar, la necesidad de climatización o el comportamiento laboral, lo que exige modelos que capturen adecuadamente estas fluctuaciones.

Muñoz et al. (2010) señalan que incorporar variables exógenas mejora la precisión de la predicción energética, especialmente cuando existen condiciones que alteran el consumo eléctrico, como variaciones en la temperatura o eventos especiales. Esta dimensión multivariada ha motivado la evolución de modelos univariados hacia estructuras más complejas como ARIMAX o VAR, capaces de capturar interdependencias dinámicas entre variables correlacionadas.

En sistemas eléctricos modernos, la predicción de demanda no solo se emplea para anticipar consumos, sino también como base para la planificación de generación, optimización del despacho energético y gestión de riesgos operativos (Weron, 2014). La precisión del pronóstico tiene un impacto directo en la eficiencia del sistema, permitiendo reducir costos y mejorar la estabilidad de la red.

En la Tabla 1, se puede observar los componentes fundamentales de una serie temporal.

**Tabla 1**

*Componentes fundamentales de una serie temporal*

Componente	Descripción
Tendencia	Variación de largo plazo en los datos, creciente o decreciente.
Estacionalidad	Patrones que se repiten en intervalos regulares (diarios, semanales).
Ciclos	Fluctuaciones asociadas a periodos irregulares de mediano o largo plazo.
Ruido	Variabilidad aleatoria no explicada por otros componentes.

*Nota.* Elaboración propia.

Si bien los modelos clásicos de series temporales han demostrado ser eficaces para capturar tendencias, estacionalidades y ciclos en la demanda eléctrica, su desempeño puede



verse limitado en contextos caracterizados por alta variabilidad, no linealidades y cambios estructurales abruptos. Estas limitaciones han motivado la incorporación progresiva de técnicas de aprendizaje automático, las cuales amplían el marco teórico tradicional al permitir la modelación de relaciones complejas sin supuestos estrictos sobre la distribución de los datos (Hastie et al., 2009).

Desde el punto de vista conceptual, esta transición no implica una sustitución de los enfoques clásicos, sino una evolución metodológica orientada a mejorar la capacidad explicativa y predictiva de los modelos aplicados a sistemas energéticos modernos (Hyndman & Athansapoulos, 2018).

### **2.2.3 Algoritmos de Machine Learning para predicción energética**

El aprendizaje automático se ha convertido en una de las áreas más prometedoras para la predicción de la demanda energética debido a su capacidad para modelar relaciones no lineales, patrones complejos y datos de alta dimensionalidad. Hastie et al. (2009) destacan que algoritmos como *random forest* y *gradient boosting* pueden manejar relaciones multivariadas complejas sin asumir estructuras lineales, lo cual es crucial en escenarios energéticos influenciados por múltiples factores.

Los modelos basados en árboles de decisión, como *random forest*, han demostrado eficacia en la predicción de cargas eléctricas por su robustez frente al ruido y su capacidad para capturar interacciones entre variables. Lago et al. (2018) evidencian que estos algoritmos superan a modelos estadísticos tradicionales cuando la demanda eléctrica presenta no linealidades y cambios abruptos en su comportamiento.

Las redes neuronales profundas han revolucionado el campo del modelado secuencial. Las redes de memoria a largo y corto plazo (long short-term memory, LSTM) introducidas por Hochreiter & Schmidhuber (1997), se posicionan como una arquitectura ideal para captar dependencias a largo plazo en series temporales energéticas. Hossain y Mahmood (2020)

muestran que las LSTM son capaces de reducir significativamente los errores de predicción en comparación con métodos clásicos, especialmente en demandas con fluctuaciones horarias intensas.

Los modelos de unidades recurrentes con compuertas (gated recurrent unit, GRU) representan una alternativa más liviana y computacionalmente eficiente que las LSTM, manteniendo resultados competitivos. (Cho et al., 2014) introducen esta arquitectura, y. Hasanat et al. (2024) reportan que las GRU son particularmente útiles en sistemas con fuerte estacionalidad, debido a su capacidad para retener información sin incrementar excesivamente el costo computacional.

Los modelos *Transformer* (Vaswani et al., 2017) han irrumpido recientemente como una herramienta de vanguardia para el análisis de secuencias. Mahmood et al. (2022) demostraron que su mecanismo de atención permite identificar relaciones globales en series temporales largas, ofreciendo ventajas frente a redes recurrentes que dependen del procesamiento secuencial.

Finalmente, la tendencia actual apunta hacia modelos híbridos o ensamblados, combinando fortalezas de modelos lineales con técnicas no lineales. Zhang (2003) fue uno de los primeros en proponer modelos híbridos ARIMA-ANN, cuyo desempeño ha sido replicado con éxito en sistemas eléctricos modernos.

En conjunto, los algoritmos de Machine Learning descritos constituyen un marco teórico robusto para el modelado de la demanda energética en sistemas eléctricos contemporáneos. Su capacidad para capturar dependencias temporales complejas, manejar grandes volúmenes de datos y adaptarse a patrones no lineales los posiciona como herramientas complementarias a los enfoques tradicionales de series temporales. Desde una perspectiva teórica, la selección y evaluación de estos algoritmos debe sustentarse en criterios de interpretabilidad, precisión y adecuación al contexto energético analizado, lo que refuerza

la necesidad de estudios comparativos que permitan identificar modelos con mayor potencial de aplicación práctica en sistemas eléctricos nacionales (Hong et al., 2016).

En la Tabla 2, se puede observar una comparación general de algoritmos aplicados a predicción energética.

**Tabla 2**

*Comparación general de algoritmos aplicados a predicción energética*

Modelo	Tipo	Ventajas	Desventajas
ARIMA/SARIMA	Estadístico	Interpretabilidad, rapidez, captura estacional	Limitado para no linealidades
Random Forest	Machine Learning	Robusto, no lineal, fácil implementación	Puede sobreajustar si no se regula
XGBoost	Machine Learning	Alto rendimiento, manejo eficiente de datos tabulares	Requiere ajuste fino de hiperparámetros
LSTM/GRU	Deep Learning	Modelan dependencias temporales complejas	Alto costo computacional
Transformer	Deep Learning	Captura dependencias globales, escalabilidad	Requiere grandes volúmenes de datos

*Nota.* Elaboración propia.

### 2.2.4 Herramientas para ciencia de datos

En este programa de estudio hemos tenido la oportunidad de trabajar a fondo varias herramientas que son imprescindibles al momento de desarrollar proyectos de ciencia de datos, pues éstas permiten realizar todos los procesos asociados al tratamiento de datos y de esa manera al fin llegar a implementar modelos de aprendizaje automático e inteligencia artificial.

Entre las herramientas más utilizadas, así como librerías podemos mencionar las siguientes:

**Python:** se trata de un lenguaje de alto nivel de código abierto creado por Guido Rossum en 1991, fu creado para simplificar la legibilidad del código en comparación con lenguajes como C++ o Java lo que simplifica su aprendizaje. Su flexibilidad permite utilizarlo para varias aplicaciones como desarrollo web, software, ciencia de datos o inteligencia artificial. Cuenta con una amplia biblioteca de librerías que facilitan las tareas complejas y disminuyen el código. (Equipo de contenidos de GoDaddy, 2025)

**Numpy:** Es una de las bibliotecas más populares de *Python* su nombre proviene de la unión de las palabras *NUMERICAL* y *PYTHON*, esta sirve para el procesamiento de vectores y matrices lo que facilita el cálculo numérico y el análisis de grandes volúmenes de datos, está dirigido a personas que trabajan en ingeniería, matemática y ciencias de datos (Ingeniería, 2024).

**Pandas:** Es una librería de código abierto para Python, flexible y robusta para trabajar con datos estructurales, se utiliza para varios procesos y es muy utilizado en ciencia de datos para realizar procesos de análisis exploratorio de datos (EDA). Se integra muy bien con otras bibliotecas como: numpy, seaborn, matplotlib, entre otras. (Almeida, 2024)

**Matplotlib:** Biblioteca de Python muy útil para crear visualizaciones de datos de manera simple, es una herramienta en la cual podemos implementar la visualización de

distintas graficas de datos para figuras como: gráfico de líneas, Histogramas o gráficos de dispersión, todos estos son importantes para tener un entendimiento grafico del sistema de datos que se está analizando (Aurora, 2024a).

**Seaborn:** Es una biblioteca de Python para visualización de datos basada en matplotlib, simplifica la creación de datos estadísticos avanzados, provee un alto nivel de personalización y una interfaz mucho más atractiva de representación de los diagramas (Aurora, 2024b).

Desde una perspectiva científica, el uso articulado de estas herramientas no responde únicamente a criterios de popularidad o disponibilidad tecnológica, sino a su capacidad para estructurar un flujo de trabajo reproducible y verificable en proyectos de ciencia de datos. Lenguajes y bibliotecas como Python, NumPy y Pandas permiten garantizar consistencia en el tratamiento de grandes volúmenes de datos, mientras que las herramientas de visualización facilitan la exploración, validación y comunicación de patrones relevantes. En el contexto de estudios energéticos, esta integración resulta fundamental para asegurar que las decisiones metodológicas estén sustentadas en datos confiables y análisis transparentes, fortaleciendo la validez de los modelos predictivos desarrollados (Mckinney,2017).

**Scikit-learn:** Librería diseñada para el desarrollo de modelos de aprendizaje automático.

**Keras:** Es una biblioteca para Python para el desarrollo de modelos de aprendizaje profundo de fácil configuración para la configuración de redes neuronales artificiales.

**Tensorflow:** Es una biblioteca de código abierto utilizada para la programación de redes neuronales y aprendizaje automático, es muy utilizada para el análisis de grafos.

**Bases de datos:** Se llama base de datos al conjunto de datos estructurados almacenada digitalmente en una unidad de memoria, pueden ser sencillas o complejas dependiendo de la

estructura y cantidad de datos que la constituyan, estas bases de datos deben ser manejadas o administradas por un sistema de gestión de bases de datos.

**API:** Una API (Application Programming Interface) es una interfaz de programación de interfaces, conocida como un conjunto de reglas o protocolos que permiten que las aplicaciones de software se comuniquen entre sí para el intercambio de datos (Michael Goodwin, n.d.).

**CSV:** Valores separados por comas, es un archivo de texto el cual separa los campos por comas realizando una especie de filas y columnas.

**JSON:** Es un formato de notación de JavaScript, es un formato basado en texto para almacenar e intercambiar datos de forma legible.

**Matriz de correlación:** Es una herramienta estadística en donde se muestra la intensidad y dirección de la relación entre dos o más variables, es muy utilizada en el campo de las finanzas, economía, psicología y biología ya que ayuda a entender cómo se relacionan las variables entre sí (Cristina Ortega, 2023).

**Machine learning:** Es el conjunto de métodos o algoritmos de la inteligencia artificial que permiten a las máquinas aprender en base a la experiencia.

**Aprendizaje supervisado:** Algoritmos que utilizan un conjunto de datos etiquetados, es decir cada dato entrenado tiene una respuesta asociada.

**Aprendizaje no supervisado:** Se refiere a algoritmos que utilizan datos que no están etiquetados de donde se intenta identificar patrones sin referencia de resultados conocidos.

**Aprendizaje por refuerzo:** Se trata de un modelo en el cual se utiliza un agente que aprende a tomar decisiones para realizar acciones para alcanzar un objetivo, basado en recompensas o penalizaciones en función de las acciones que realiza para alcanzar el objetivo.

**Aprendizaje profundo:** Es un subconjunto del machine learning impulsado por redes neuronales, cuyo diseño está diseñado inspirado en la estructura del cerebro humano (Cole Stryker, 2025).

**Redes Neuronales artificiales:** Es un modelo de aprendizaje automático en el cual se apilan neuronas simples en capas, estas últimas aprenden pesos y sesgos para el reconocimiento de patrones de datos que se asignan a las salidas (Cole Stryker, 2025).

**Red Neuronal recurrente:** red neuronal profunda que se entrenan con datos secuenciales o series temporales para crear un modelo de ML para predicciones secuenciales.

**Series temporales:** Conjunto de datos ordenados en el tiempo por lo general se ordena en periodos de tiempo regulares.

**Hiperparámetros:** Son variables de parametrización para modelos de ML que se establecen antes de un entrenamiento para controlar su proceso de aprendizaje.

En el ámbito de la predicción de la demanda energética, estas categorías de aprendizaje automático adquieren relevancia en función de su capacidad para modelar relaciones complejas entre variables históricas y exógenas. Los enfoques supervisados han sido ampliamente utilizados para estimar consumos futuros a partir de datos etiquetados, mientras que las técnicas no supervisadas permiten identificar patrones latentes y segmentaciones de comportamiento en los perfiles de consumo. Por su parte, los modelos basados en redes neuronales recurrentes y arquitecturas profundas han demostrado un desempeño superior en la captura de dependencias temporales de largo plazo, lo que resulta especialmente pertinente en sistemas eléctricos con alta estacionalidad y variabilidad climática. Esta adecuación teórica justifica su inclusión como base conceptual del presente estudio (Goodfellow et al, 2016).

**Métricas de desempeño:** son herramientas que permiten evaluar la eficacia de los algoritmos frente al procesamiento de datos, presentan una medición cuantitativa para guiar a los científicos de datos para determinar o ajustar sus modelos a los objetivos planteados.

**MAE:** Se lo conoce como el Error Absoluto Medio, que se calcula como la diferencia de absoluta de los valores reales y los valores predichos, sirve para evidenciar una desviación general en la predicción sin dar importancia a errores grandes.

**RMSE:** Conocido como Raíz del Error cuadrático Medio, que se determina con la raíz cuadrada de la diferencia de los errores al cuadrado, es muy útil cuando se quiere identificar errores grandes.

**MAPE:** Es el Error Porcentual Absoluto Medio, se realiza con los errores como porcentajes, ideal para comparar conjuntos de datos, pero tiene una limitación cuando el error es cercano a cero.

**Streamlit:** Es una biblioteca de Python de código abierto para poder crear aplicaciones webs interactivas para proyectos de ciencia de datos y machine learning.

**Facebook Prophet:** es una Biblioteca para Python creada por Facebook adecuada para realizar pronóstico de series temporales de forma sencilla.

**XGBoost:** Es una biblioteca de código abierto para aprendizaje automático para problemas de regresión que combina muchos arboles de decisión en su algoritmo para formar un modelo robusto.

En conclusión, las herramientas para la ciencia de datos analizadas en este subtema constituyen un soporte esencial para el desarrollo riguroso de procesos analíticos orientados a la modelación y predicción en contextos complejos. Su valor no radica únicamente en su funcionalidad técnica, sino en su capacidad para integrarse de manera coherente dentro de un flujo metodológico estructurado, que garantice la calidad, consistencia y trazabilidad de los datos a lo largo de todo el proceso investigativo. Lenguajes de programación, bibliotecas



especializadas, algoritmos de aprendizaje automático y métricas de evaluación conforman un ecosistema que posibilita la exploración, transformación y análisis de grandes volúmenes de información de forma sistemática y reproducible.

Finalmente, la adecuada selección y articulación de estas herramientas permite fortalecer la validez científica de los modelos predictivos, al facilitar la identificación de patrones relevantes, la evaluación objetiva del desempeño y la reducción de sesgos en la toma de decisiones. En el ámbito de la predicción energética, estas capacidades adquieren especial relevancia debido a la naturaleza dinámica y multivariable de los sistemas eléctricos, donde la precisión y confiabilidad de los resultados dependen directamente del manejo adecuado de los datos y de la correcta implementación de los métodos analíticos.

### ***2.2.5 Metodologías de ciencia de datos***

El ciclo de vida de la inteligencia artificial constituye una de las vías más efectivas para incrementar las probabilidades de éxito en un proyecto de ciencia de datos, ya que implica el seguimiento sistemático de un proceso estructurado hasta alcanzar la solución planteada. Este ciclo de vida, también denominado metodología de ciencia de datos, describe un enfoque secuencial que orienta la ejecución del proyecto y proporciona una guía clara para que los científicos de datos comprendan de manera intuitiva cada una de sus etapas.

La aplicación de una metodología resulta fundamental en cualquier campo del conocimiento, particularmente en la ciencia de datos, debido a varias razones esenciales. En primer lugar, permite garantizar la coherencia de los resultados, incrementando la probabilidad de obtener conclusiones consistentes y predecibles. Asimismo, ofrece una orientación clara para profesionales o ingenieros con menor experiencia, facilitando su integración en proyectos complejos. Finalmente, contribuye a establecer supuestos, expectativas y un vocabulario común, lo que favorece la alineación de objetivos y una interacción efectiva entre los miembros del equipo de trabajo.

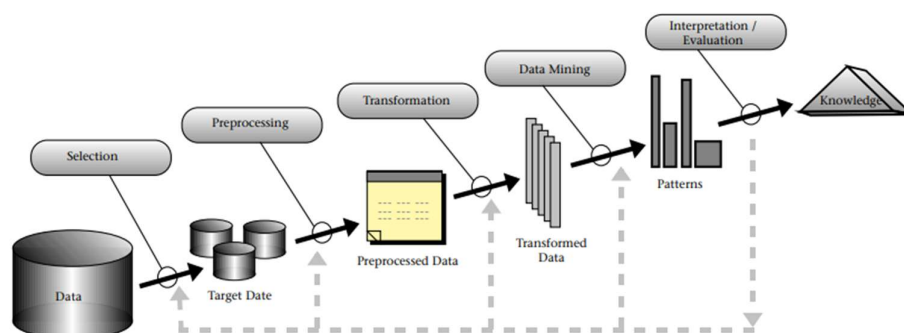
A continuación, se muestra un resumen de las metodologías más conocidas de donde saldrá la que se adapte a nuestra necesidad para nuestro proyecto en estudio.

La metodología KDD (Knowledge Discovery in Databases) constituye un proceso de uso extendido en los campos de la minería de datos y el aprendizaje automático, cuyo propósito central es extraer conocimiento valioso a partir de grandes cantidades de información. Este enfoque se desarrolla a través de una serie de etapas estrechamente vinculadas entre sí. En primer lugar, se lleva a cabo la selección de las fuentes y muestras de datos que serán analizadas. Después, el preprocesamiento se encarga de depurar y normalizar los registros para prevenir dificultades en las fases siguientes. Durante la transformación, se recurre a técnicas de reducción de dimensionalidad que permiten descartar información irrelevante y simplificar el conjunto de datos. Posteriormente, la minería de datos consiste en aplicar algoritmos de aprendizaje automático para identificar patrones y relaciones de interés. Por último, la interpretación posibilita el análisis de los hallazgos a través de visualizaciones claras, lo cual contribuye a fundamentar la toma de decisiones (Admin, 2020).

El proceso KDD permiten realizar un procesamiento estructurado el mismo que puede ser refinado en función de resultados obtenidos en etapas posteriores, este además brinda un procesamiento exhaustivo de los datos, el mismo que se puede aplicar en varios dominios o campos de estudio para ayudar a la toma de decisiones informadas (Oleh Dubetcky, 2024).

**Figura 2**

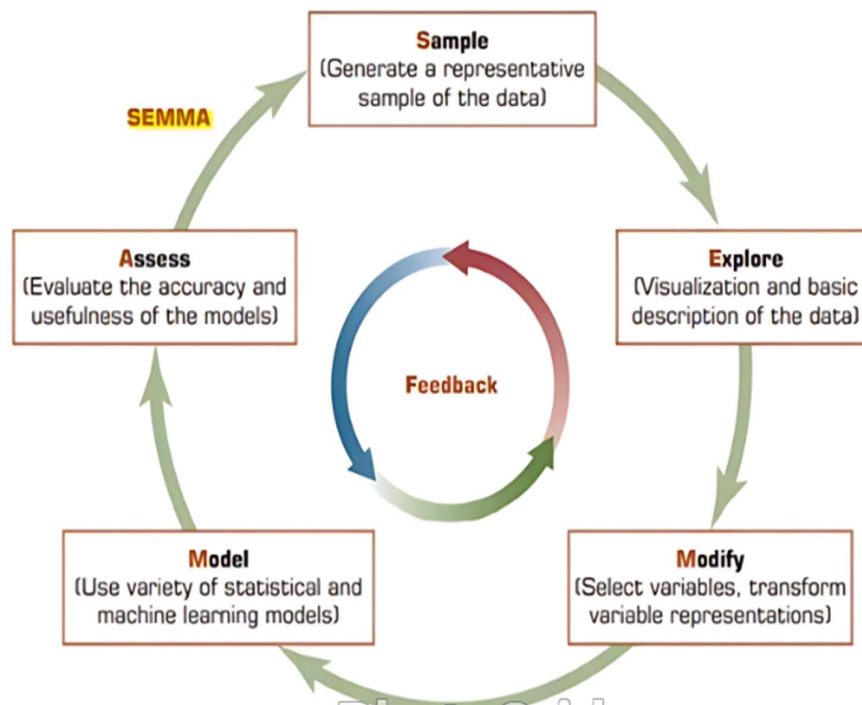
*Descripción general de los pasos que componen el proceso KDD*



*Nota.* El gráfico representa los pasos que componen el proceso KDD (Fayyad et al., 1996).

En cuanto a la metodología SEMMA (Sample, Explore, Modify, Model and Assess), propuesta por el SAS Institute, esta se enfoca en el análisis y la construcción de modelos predictivos dentro del campo de la minería de datos, siguiendo un esquema práctico dividido en cinco etapas. El proceso comienza con el muestreo, donde se extrae una porción de datos que sea representativa del conjunto total. Luego, en la exploración, se examinan los datos con el fin de reconocer tendencias y patrones que orienten el análisis posterior. La modificación implica depurar y ajustar los datos para que resulten aptos para su procesamiento. En el modelado se emplean técnicas como redes neuronales o árboles de decisión para generar predicciones, y finalmente, la evaluación permite determinar si los resultados alcanzados son precisos y confiables (Admin, 2020).

Esta metodología es ampliamente utilizada en el ámbito industrial por su enfoque práctico y confiable (Oleh Dubetcky, 2024).

**Figura 3***Modelo SEMMA*

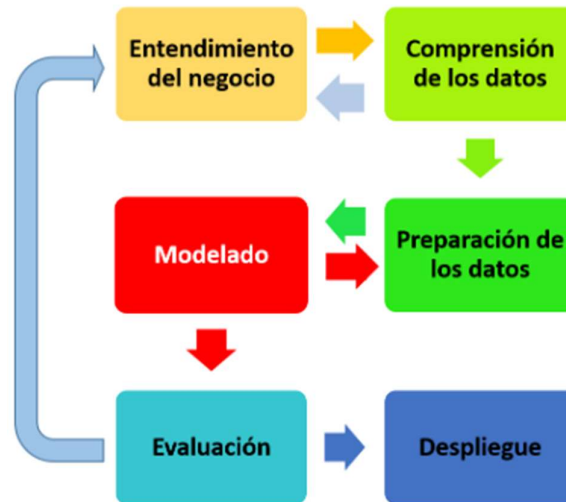
*Nota.* Modelo SEMMA (Sharda, 2018)

Finalmente, la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) surge como una propuesta más completa que las anteriores, distinguiéndose por el rigor con el que documenta cada etapa del proceso analítico. Su estructura contempla seis fases interconectadas: inicia con la comprensión del negocio, etapa en la cual se establecen los propósitos que guiarán el análisis; posteriormente, la comprensión de los datos facilita un acercamiento exploratorio que busca detectar patrones iniciales y plantear hipótesis de trabajo. La preparación de los datos abarca actividades como la selección, depuración y transformación de la información disponible. Durante el modelado se implementan técnicas de aprendizaje automático que respondan a las hipótesis formuladas, y en la evaluación se contrasta la validez y fiabilidad de los hallazgos obtenidos. El ciclo culmina con el

despliegue, fase dedicada a materializar el producto final o integrarlo en un ambiente operativo (Admin, 2020).

**Figura 4**

*Etapas de metodología CRISP-DM*



*Nota.* Etapas de metodología CRISP-DM (Fernández-Avilés, 2024)

Así como otras metodologías su aplicación iterativa permite ir ajustando de mejor manera los resultados en base a los conocimientos que se van obteniendo en los procesos intermedios, esto hace da flexibilidad de aplicación y es muy popular en la comunidad de ciencia de datos. (Oleh Dubetcky, 2024)

**Tabla 3***Comparación de metodologías de ciencia de datos*

<b>Aspecto</b>	<b>KDD</b>	<b>SEMMA</b>	<b>CRISP-DM</b>
<b>Origen</b>	Comunidades académicas y de investigación	Instituto SAS	Iniciativa intersectorial
<b>Enfoque</b>	Proceso de descubrimiento de conocimiento	Procesos técnicos de minería de datos	Metodología estructurada de minería de datos
<b>Fases</b>	Selección, Preprocesamiento, Transformación, Minería de Datos, Interpretación/Evaluación	Muestreo, Exploración, Modificación, Modelado, Evaluación	Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue
<b>Flexibilidad</b>	Alta	Moderada	Alta

*Nota.* Elaboración propia.

### 3. CAPÍTULO 3: DESARROLLO

#### 3.1 Metodología

Para el desarrollo de nuestro proyecto se ha visto conveniente utilizar la metodología KDD ya que se ajusta a nuestro objetivo de estudio para el descubrimiento de patrones de comportamiento de consumo de energía eléctrica en Ecuador, por ello vamos a seguir las etapas recomendadas en función de las necesidades que se van presentando a lo largo del desarrollo del proyecto.

Además, la metodología KDD permite estructurar el proceso de análisis de datos de manera iterativa y flexible, lo cual resulta especialmente relevante en proyectos donde la calidad y la disponibilidad de los datos pueden variar en el tiempo. En el contexto del consumo energético, esta característica es clave, ya que los registros históricos suelen estar sujetos a inconsistencias, valores faltantes o cambios en los criterios de medición.

Al apoyarse en fases bien definidas como la selección, limpieza, transformación y modelado de los datos, KDD facilita no solo la obtención de modelos predictivos más robustos, sino también una comprensión más profunda de los patrones subyacentes en el comportamiento del consumo eléctrico. Este enfoque metodológico ha sido ampliamente adoptado en proyectos de analítica energética debido a su capacidad para integrar técnicas estadísticas y de aprendizaje automático dentro de un marco sistemático y reproducible.

Así mismo hemos denotado de un nombre a nuestra aplicación de predicción eléctrica como ELFO (Electric Forecast), nace de la necesidad de modernizar y optimizar la planificación energética en Ecuador. Se trata de una solución web integral diseñada para abordar el ciclo completo de predicción de demanda: desde la ingesta de datos crudos operativos hasta la generación de pronósticos certeros mediante Inteligencia Artificial.

La elección de una solución web como plataforma para ELFO responde también a la necesidad de democratizar el acceso a herramientas de análisis y predicción energética.

Tradicionalmente, este tipo de análisis ha estado limitado a entornos técnicos especializados, lo que dificulta su uso por parte de planificadores, analistas y tomadores de decisiones no técnicos. Al centralizar el procesamiento y la visualización de resultados en una aplicación web, el sistema facilita la interpretación de la información y promueve una toma de decisiones más informada y oportuna. Este enfoque se alinea con las tendencias actuales en sistemas de apoyo a la decisión, donde la usabilidad y la claridad en la presentación de resultados son tan importantes como la precisión de los modelos predictivos.

A diferencia de soluciones manuales propensas a error, ELFO propone una arquitectura automatizada y reproducible. Aunque es un prototipo académico, ha sido construido siguiendo buenas prácticas de ingeniería de software y ciencia de datos, buscando un equilibrio entre la complejidad teórica y la usabilidad práctica.

En este sentido, la automatización del flujo de trabajo no solo reduce la probabilidad de errores humanos, sino que también permite la trazabilidad completa del proceso de análisis, desde la ingesta de los datos hasta la generación de los resultados finales. Esta trazabilidad es un aspecto fundamental en proyectos que trabajan con datos críticos, como los relacionados con la planificación energética nacional, ya que facilita la auditoría, validación y replicabilidad de los resultados obtenidos. Asimismo, el uso de buenas prácticas de ingeniería de software contribuye a que el sistema pueda evolucionar en el tiempo, incorporando nuevas fuentes de datos o modelos predictivos sin comprometer su estabilidad ni su desempeño general.

### **Arquitectura del Sistema**

Para garantizar mantenibilidad y escalabilidad, se optó por una arquitectura monolítica pero altamente modular. La separación de responsabilidades es clara, lo que facilita futuras auditorías o mejoras en componentes específicos sin romper el sistema general.



La adopción de una arquitectura modelar también favorece la escalabilidad del sistema frente a incrementos en el volumen de datos o en la complejidad de los análisis realizados. En escenarios reales, los sistemas de predicción de demanda energética deben ser capaces de adaptarse a horizontes temporales más amplios y a una mayor granularidad en los datos, como registros horarios o incluso en tiempo real casi real.

**Frontend (Interfaz de Usuario):** Desarrollado en **Streamlit**. Se eligió por su capacidad de prototipado rápido y su ecosistema nativo de datos, permitiendo visualizar gráficos complejos con pocas líneas de código.

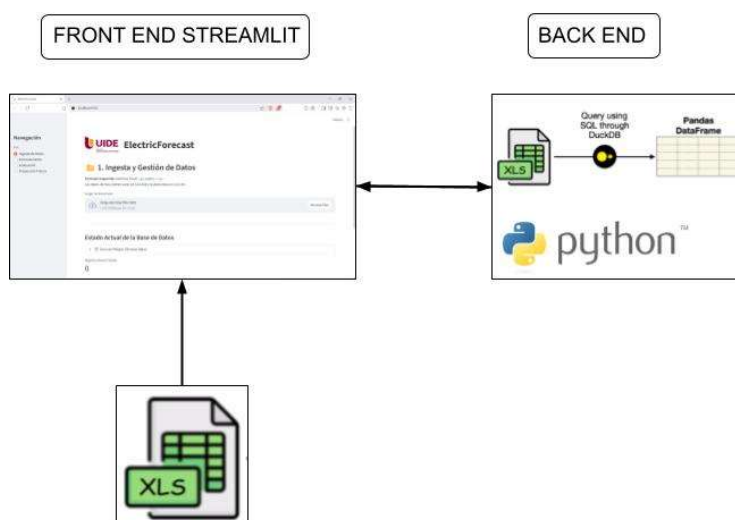
**Backend (Lógica de Negocio):** Implementado en **Python 3.10+**, aprovechando su dominio en el campo del Machine Learning.

**Procesamiento ETL:** Uso intensivo de **Pandas** y **NumPy** para manipulación vectorial de datos.

**Modelado Predictivo:** Integración híbrida de enfoques estadísticos (**Facebook Prophet**) y de Machine Learning clásico (**XGBoost**).

**Figura 5**

*Arquitectura y flujo de información.*



*Nota.* Elaboración propia.

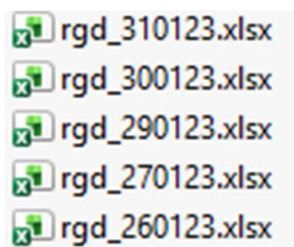
### 3.1.1 Selección de Datos

Para esta sección de selección de datos (*ingesta de datos*), se utilizó bases de datos de reportes de consumo energético obtenidos de la corporación nacional de electricidad CNEL, las cuales se encuentran distribuidas por lotes de acuerdo con los meses de consumo de cada año que se toma como referencia para nuestro estudio. Estos datos vienen en formato *Excel* con los detalles de consumo hora a hora, por lo tanto, estos documentos planos debemos consolidarlos en una base de datos que nos permita procesarlos adecuadamente en los pasos posteriores.

La selección de los reportes de consumo energético como fuente principal de datos responde a su carácter oficial y a su consistencia histórica, lo cual resulta fundamental para estudios orientados a la predicción de demanda. Al provenir de una entidad nacional, estos registros ofrecen un nivel adecuado de confiabilidad y representatividad del comportamiento real del sistema eléctrico. Asimismo, el hecho de que los datos estén organizados por periodos mensuales facilita el análisis temporal y permite evaluar variaciones estacionales en el consumo, aspecto clave en modelos de series temporales aplicados al sector energético.

#### Figura 6

*Muestra de reporte diarios de consumo*



*Nota.* Elaboración propia.

Como se observa en la figura anterior los reportes disponibles tienen una nomenclatura con el patrón “rgd\_DDMMYY.xlsx”, por tanto, el sistema se diseñó para consumir este tipo de documentos.

La estandarización en la nomenclatura de los archivos no solo simplifica el proceso de ingestión automática de datos, sino que también reduce la probabilidad de errores humanos durante la carga y validación de la información. Este tipo de convenciones resulta especialmente relevante en sistemas que procesan grandes volúmenes de archivos de manera recurrente, ya que permite identificar de forma inequívoca el periodo temporal asociado a cada conjunto de datos.

Nuestro procesamiento de datos damos especial importancia al proceso ETL ya que sabemos que “basura entra, basura sale” en ese sentido se dedica un esfuerzo considerable a la limpieza de datos y considerando nuestra arquitectura modular se crea un módulo (etl\_processor.py) para manejar este proceso de forma adecuada.

### Figura 7

*Tabla de reporte de producción de energía nacional.*

HORA	GENERACION							TIE		TOTAL SISTEMA	DEMANDA EMPRESAS ELECTRICAS
	HIDRAULICA EMBALSE	HIDRAULICA PASADA	GAS	VAPOR	BUNKER	DIESEL	NO CONVENCIONAL	IMPORTACION COLOMBIA	IMPORTACION PERU		
	MW	MW	MW	MW	MW	MW	MW	MWh	MWh		
01:00	927.4	2448.4	55.5	52.5	55.2	-0.2	82.3	-2.6	0.0	3618.6	3408.2
02:00	799.1	2428.4	58.8	52.3	23.2	-0.1	86.5	0.0	0.0	3448.1	3236.8
03:00	741.3	2385.1	60.6	52.0	22.4	-0.1	80.4	-8.5	0.0	3333.1	3145.4
04:00	749.8	2264.0	60.4	52.5	23.4	-0.1	85.5	0.0	0.0	3235.6	3043.8
05:00	697.3	2253.3	60.6	52.4	23.0	-0.1	90.5	0.0	0.0	3177.0	2989.9
06:00	720.7	2184.6	60.9	52.7	22.3	-0.1	77.3	26.0	0.0	3144.3	2955.1
07:00	701.6	2093.9	59.9	52.6	22.9	3.6	89.3	14.0	0.0	3038.0	2876.4
08:00	467.2	2138.2	60.1	52.4	42.7	3.6	91.0	5.4	0.0	2860.5	2688.0
09:00	641.3	2134.6	59.8	52.3	45.3	3.6	91.5	-19.7	0.0	3008.7	2837.4
10:00	805.2	2188.7	58.9	52.3	43.7	3.6	93.6	-2.6	0.0	3243.3	3053.1
11:00	816.0	2328.2	58.9	52.5	44.4	3.5	107.9	11.0	0.0	3422.4	3219.2
12:00	893.5	2383.6	60.4	53.1	54.4	4.1	146.0	14.9	0.0	3610.0	3364.4
13:00	989.1	2405.8	59.1	52.7	66.3	4.0	151.2	13.8	0.0	3742.1	3497.4
14:00	1048.3	2439.5	55.3	52.9	79.5	4.1	146.0	0.0	0.0	3825.7	3572.8
15:00	1063.1	2495.4	57.7	52.8	80.3	4.1	140.7	0.0	0.0	3894.2	3650.5
16:00	1087.3	2520.3	56.0	52.8	81.2	4.2	118.6	0.0	0.0	3920.4	3685.8
17:00	1069.9	2474.4	54.8	93.2	128.9	0.5	115.2	9.6	0.0	3946.4	3706.4
18:00	1070.0	2306.9	55.5	93.0	203.8	46.6	91.0	11.0	0.0	3877.7	3673.4
19:00	1054.3	2281.9	56.6	123.4	237.3	68.4	85.5	10.2	0.0	3917.7	3749.0
20:00	1138.7	2671.3	57.5	137.6	246.4	70.3	81.5	0.0	0.0	4403.4	4234.9
21:00	1092.1	2689.4	56.5	137.3	255.0	70.0	80.1	13.3	0.0	4393.7	4218.7
22:00	1085.6	2687.4	55.8	137.1	252.8	70.9	74.0	-26.3	0.0	4337.3	4184.9
23:00	1027.8	2582.2	58.1	137.1	267.7	66.9	74.0	3.5	0.0	4217.2	4049.5
00:00	952.3	2414.9	55.9	137.1	268.3	70.7	73.7	-23.5	0.0	3949.3	3810.0
01:00	806.7	2259.7	55.9	136.9	258.7	70.0	73.7	12.0	0.0	3673.7	3532.7
02:00	707.6	2132.6	58.8	137.7	260.4	70.0	81.1	0.0	0.0	3448.1	3305.1

*Nota.* Elaboración propia.

### 3.1.2 *Preprocesamiento de Datos*

Como ya se entendió la organización de los reportes que van a alimentar nuestro aplicativo, continuamos con el preprocesamiento de la data cruda para ir la adecuando a nuestro requerimiento y poder realizar el modelado futuro.

El preprocesamiento constituye una de las etapas más críticas dentro del flujo de análisis de datos, ya que la calidad de los modelos predictivos depende en gran medida de la calidad de la información utilizada durante su entrenamiento. En el caso de datos energéticos, es común encontrar inconsistencias derivadas de errores de medición, interrupciones en el registro o cambios operativos en el sistema eléctrico. Por esta razón, se adoptó un enfoque riguroso de preparación de datos que prioriza la coherencia temporal y la eliminación de ruido, garantizando que los modelos aprendan patrones reales y no artefactos del proceso de recolección.

**Validación de Formato:** Se implementaron expresiones regulares (Regex) para extraer y validar la fecha directamente del nombre del archivo, asegurando que el contenido corresponda al día reportado.

La validación temprana de los datos permite detectar inconsistencias antes de que estas se propaguen a etapas posteriores del procesamiento. Al verificar que la fecha extraída del nombre del archivo coincida con el contenido del reporte, se asegura la integridad temporal de los registros y se previenen errores que podrían afectar el entrenamiento de los modelos predictivos.

Este tipo de controles resulta especialmente importante en aplicaciones automatizadas, donde los errores no detectados pueden escalar rápidamente y comprometer la confiabilidad de los resultados finales.

## Figura 8

### *Extracción y validación de fechas*

```
def parse_date_from_filename(filename):
    """
    Extracts date from filename format 'rgd_DDMMYY.xlsx'.
    """
    # Regex to match 'rgd_DDMMYY'
    match = re.search(r'rgd_(\d{6})', filename, re.IGNORECASE)
    if match:
        date_str = match.group(1)
        try:
            # Parse DDMMYY
            date = pd.to_datetime(date_str, format='%d%m%y')
            return date
        except ValueError:
            return None
    return None
```

*Nota.* Elaboración propia.

**Lectura de Rango Específico:** Se extraen los datos del rango A12:L36. Aquí nos enfrentamos a un reto particular: la representación de las 24 horas. El sistema maneja la conversión de la hora "24:00" como las "00:00" del día siguiente, una peculiaridad común en sistemas eléctricos que puede romper librerías estándar de tiempo si no se trata manualmente.

La selección de un rango específico de celdas permite aislar la información relevante y evitar la inclusión de encabezados, totales u otros elementos que no forman parte del análisis cuantitativo. Este enfoque contribuye a una estructura de datos más limpia y facilita la posterior transformación a formatos compatibles con bibliotecas de análisis en Python.

### **3.1.3 Transformación**

Antes de entrenar cualquier modelo, los datos pasan por un "lavado" riguroso a nivel horario.

La fase de transformación tiene como objetivo principal adaptar los datos a un formato que maximice la capacidad de aprendizaje de los modelos predictivos. En esta etapa, se busca no solo corregir errores o inconsistencias, sino también resaltar las características relevantes del comportamiento del consumo eléctrico. Una transformación adecuada permite

reducir la complejidad innecesaria de los datos y mejorar la eficiencia computacional del entrenamiento, lo cual es especialmente importante cuando se trabaja con grandes volúmenes de información histórica.

**Detección de Anomalías (Outliers) con Z-Score:** Para identificar valores atípicos (picos de demanda irreales por errores de medición), calculamos el **Z-Score** de cada punto  $Z = \frac{x-\mu}{\sigma}$  donde  $\mu$  representa la media y  $\sigma$  la desviación estándar del conjunto de datos, de acuerdo con la definición clásica del estadístico Z en estadística inferencial (Walpole et al., 2012).

Se definió un umbral de  $|Z| > 3$ . Esto significa que cualquier consumo que se desvíe más de 3 desviaciones estándar del promedio histórico es marcado inmediatamente como un error. Este es un método estadístico robusto y simple que asume una distribución normal de los errores.

La identificación de valores atípicos permite distinguir entre variaciones normales del consumo y registros anómalos que no representan el comportamiento real del sistema eléctrico. En contextos energéticos, estos valores pueden surgir por fallos en los sistemas de medición, interrupciones del servicio o eventos excepcionales. El uso del Z-Score ofrece una solución sencilla y efectiva para detectar este tipo de anomalías, permitiendo tratarlas de manera controlada sin eliminar información potencialmente valiosa para el análisis.

## Figura 9

### *Cálculo de Z-score*

```
# Z-Score
mean = df_clean['valor'].mean()
std = df_clean['valor'].std()

if std == 0:
    df_clean['is_outlier'] = False
else:
    df_clean['z_score'] = (df_clean['valor'] - mean) / std
    df_clean['is_outlier'] = df_clean['z_score'].abs() > 3
```

*Nota.* Elaboración propia.

**Manejo de Valores Nulos:** Tanto los datos faltantes originales como los outliers detectados se enmascaran como NaN (Not a Number).

## Figura 10

### *Manejo de valores nulos*

```
# Mask Outliers and Nulls as NaN for interpolation
mask = df_clean['is_outlier'] | df_clean['is_null']
df_clean.loc[mask, 'valor'] = np.nan
```

*Nota.* Elaboración propia.

**Imputación por Interpolación Lineal:** Dado que la demanda eléctrica es una variable continua y física (no cambia de 0 a 100 en un segundo), la mejor estrategia de imputación es la interpolación lineal.

Rellenamos los huecos trazando una línea recta entre los puntos válidos adyacentes. Esto preserva la tendencia local mejor que rellenar con un promedio global.

**Figura 11***Interpolación lineal*

```
# Interpolate  
df_clean['valor'] = df_clean['valor'].interpolate(method='linear', limit_direction='both')
```

*Nota.* Elaboración propia.

**Agregación Temporal:** Finalmente, aunque la ingesta es horaria, el modelado se realiza a nivel **DIARIO**. Se aplica un *resampling* promediando los valores horarios limpios. Esto suaviza el ruido de alta frecuencia y permite a los modelos enfocarse en las tendencias estacionales macro.

La agregación temporal a nivel diario responde a la necesidad de equilibrar el nivel de detalle de los datos con la estabilidad de las predicciones. Al reducir la frecuencia horaria, se atenúan fluctuaciones de corto plazo que pueden introducir ruido en el entrenamiento de los modelos, permitiendo que estos se concentren en tendencias más representativas del comportamiento general de la demanda.



**Figura 12***Resampling temporal*

```

# Clean Hourly First
df = df.copy()
df['fecha'] = pd.to_datetime(df['fecha'])
df_clean_hourly, df_dirty_hourly, stats_hourly = detect_outliers_and_impute(df.sort_values('fecha'))

# Resample to Daily
# Dirty Daily (just mean of raw, with nans propagating)
daily_dirty = df_dirty_hourly.set_index('fecha').resample('D')['valor'].mean().reset_index()

# Clean Daily (mean of clean hourly)
daily_clean = df_clean_hourly.set_index('fecha')['valor'].resample('D').mean()

# But wait, if we have missing HOURS in a day, the daily mean might be skewed if we don't fill holes first.
# We already filled holes in df_clean_hourly.

# Also check for MISSING DAYS (gaps in files).
# Reindex daily to full range
full_idx = pd.date_range(start=daily_clean.index.min(), end=daily_clean.index.max(), freq='D')
daily_clean = daily_clean.reindex(full_idx)

# Interpolate missing days if any (file gaps)
daily_clean_imputed = daily_clean.interpolate(method='linear')

```

*Nota.* Elaboración propia.

**3.1.4 Minería de datos**

Se implementó un enfoque de "competencia de modelos", entrenando dos arquitecturas fundamentalmente distintas para permitir al usuario comparar resultados.

**Prophet (Enfoque Estadístico Aditivo).**

Utilizamos este enfoque desarrollado por META por su capacidad nativa para manejar series temporales con fuertes componentes estacionales y efectos de días festivos, algo crítico en el consumo eléctrico por el comportamiento de la población y producción en días decretados como descanso nacional.

**Componentes:** Se desactivó la estacionalidad diaria (ya que nuestros datos son diarios) pero se forzó la estacionalidad semanal y anual.

**Localización:** Se integró el calendario de feriados de Ecuador (`add_country_holidays('EC')`). Esto es vital, ya que la demanda cae drásticamente en feriados; un modelo que no sepa esto fallará sistemáticamente en esas fechas.

**Figura 13**

*Configuración de feriados en modelo Prophet*

```
class ProphetModel:
    def __init__(self, holidays=True):
        self.holidays = holidays
        self.model = None
```

*Nota.* El código describe, toma en cuenta los días de feriado.

### **XGBoost con Optimización Bayesiana (Machine Learning).**

Aquí es donde el proyecto profundiza técnicamente. XGBoost (Extreme Gradient Boosting) es un algoritmo de ensamble de árboles de decisión, famoso por ganar competencias de Kaggle.

**Figura 14**

*Configuración de modelo XGBoost*

```
class XGBoostModel:
    def __init__(self, use_bayesian_opt=False):
        self.model = None
        self.feature_names = []
        self.use_bayesian_opt = use_bayesian_opt
        self.best_params = {}
```

*Nota.* El código buscará automáticamente los mejores parámetros para funcionar mejor, usando una técnica inteligente llamada optimización bayesiana.

**Ingeniería de Características (Feature Engineering):** XGBoost no "entiende" el tiempo como una secuencia. Por ello, transformamos la fecha en características numéricas explícitas:

Día de la semana (0-6)

Trimestre y Mes

Día del año Esto permite al árbol "cortar" los datos y aprender patrones como "los lunes son más altos" o "diciembre tiene picos".

### Figura 15

*Ingeniería de características*

```
def create_features(self, df):
    df = df.copy()
    df['dayofweek'] = df['fecha'].dt.dayofweek
    df['quarter'] = df['fecha'].dt.quarter
    df['month'] = df['fecha'].dt.month
    df['year'] = df['fecha'].dt.year
    df['dayofyear'] = df['fecha'].dt.dayofyear
    return df

def train(self, df):
    df_feats = self.create_features(df)

    X = df_feats.drop(columns=['fecha', 'valor'])
    y = df_feats['valor']
```

*Nota.* Elaboración propia.

**Validación Cruzada Temporal (Time Series Split):** Para validar el modelo, no usamos un `train_test_split` aleatorio (que sería un error grave en series temporales). Usamos

TimeSeriesSplit, el cual entrena el conjunto de datos en el pasado y prueba en el futuro de forma secuencial respetando la causalidad temporal.

### Optimización de Hiperparámetros: Búsqueda Bayesiana

En lugar de usar un "Grid Search" (que prueba todas las combinaciones a fuerza bruta y es muy lento) o un "Random Search" (que es suerte), implementamos una Búsqueda Bayesiana (BayesSearchCV de scikitlearn).

¿Por qué Bayesiana? Este método es "inteligente". Construye un modelo de probabilidad de la función objetivo (en nuestro caso, minimizar el error cuadrático). Aprende de las pruebas anteriores. Si probar un learning\_rate=0.01 dio mal resultado, el algoritmo “sabe” no buscar cerca de ahí y explora áreas más prometedoras.}

#### Figura 16

*Optimización de búsqueda bayesiana*

```
self.use_bayesian_opt:
    print("Iniciando Búsqueda Bayesiana de Hiperparámetros...")
    # Define search space
    search_space = {
        'max_depth': Integer(3, 10),
        'learning_rate': Real(0.01, 0.3, prior='log-uniform'),
        'n_estimators': Integer(100, 1000),
        'subsample': Real(0.5, 1.0),
        'colsample_bytree': Real(0.5, 1.0)
    }
```

*Nota.* Elaboración propia.

#### Espacio de Búsqueda Definido.

Configuramos la búsqueda para explorar rangos dinámicos:

max\_depth (3-10): Para controlar la complejidad de los árboles y evitar sobreajuste.

learning\_rate (0.01 - 0.3): Usando una distribución log-uniforme para probar con más detalle valores pequeños (que suelen ser mejores).

subsample y colsample\_bytree: Para agregar aleatoriedad y robustez al ensamble.

El resultado es un modelo XGBoost "tuneado" específicamente para los datos de Ecuador, superando significativamente a una configuración por defecto.

## 4. CAPÍTULO 4: ANÁLISIS DE RESULTADOS

En este capítulo abordaremos el procesamiento propiamente de los reportes de generación disponibles desde el año 2022, los cuales se presentan como hojas de cálculo individuales para cada día del año, a continuación, realizamos los ejercicios de predicción en nuestra plataforma web “ELFO” de pruebas, que se aloja localmente (localhost) en el computador que corre en una estación de trabajo.

Es importante destacar que la utilización de un entorno de pruebas local permitió un mayor control sobre el flujo de ejecución del sistema, reduciendo variables externas que podrían afectar los resultados del análisis. Esta decisión facilitó la validación del correcto funcionamiento de cada etapa del procesamiento, desde la carga de los archivos hasta la generación de las predicciones. Además, el uso de datos históricos consolidados a partir del año 2022 proporcionó una base temporal suficiente para evaluar el comportamiento del consumo eléctrico en distintos escenarios, permitiendo identificar patrones recurrentes y variaciones estacionales relevantes para el análisis predictivo.

### 4.1 Pruebas de concepto

Nuestras pruebas de concepto se centran en la aplicación del proceso ETL para procesamiento de datos, proceso EDA para el análisis e imputación de registros faltantes, finalmente se realiza el entrenamiento con los dos modelos machine learning escogidos para series temporales, Prophet y XGBoost.

Las pruebas de concepto realizadas cumplen un rol fundamental dentro del desarrollo del proyecto, ya que permiten verificar, en un entorno controlado, la viabilidad técnica de las soluciones propuestas antes de su aplicación definitiva. En este contexto, la correcta ejecución del proceso ETL y del análisis exploratorio de datos garantiza que los modelos predictivos se entrenen con información consistente y representativa del fenómeno estudiado.

Nuestro dataset fue procesado de la consolidación de los reportes individuales que se alimentan desde el panel de ingesta de datos, adicional se realiza la transformación del formato de fecha y se extrae la columna de interés donde está el consumo general de electricidad a nivel nacional.

La consolidación de los reportes individuales en un único conjunto de datos permitió obtener una visión integral del consumo eléctrico nacional, reduciendo la fragmentación de la información y facilitando el análisis temporal. Este enfoque es especialmente relevante en estudios energéticos, donde la agregación adecuada de los datos contribuye a disminuir el ruido y resaltar tendencias de fondo.

Para la ejecución de las predicciones en nuestra plataforma ELFO se ha establecido 4 niveles de procesamiento los cuales son muy intuitivos para facilidad de uso del sistema, estos niveles de navegación son: ingesta de datos, entrenamiento, evaluación y proyección futura.

#### **4.2 Ingesta de datos en ELFO**

La interfaz permite la ingesta de datos en su respectiva sección para lo cual hay que dirigirse al botón de “Browse files” o a su vez también se lo puede arrastrar los documentos directamente desde un explorador de archivos.

La flexibilidad en la carga de archivos mejora significativamente la experiencia del usuario y reduce la probabilidad de errores durante la ingesta de datos. Al permitir tanto la selección manual como el arrastre directo de archivos, el sistema se adapta a distintos perfiles de usuario y flujos de trabajo. Esta característica resulta especialmente útil en entornos institucionales, donde los datos suelen generarse y actualizarse de forma periódica.

**Figura 17***Carga de datos en ELFO*

The screenshot displays the ElectricForecast web application interface. On the left, a OneDrive file explorer shows a directory structure under 'UIDE > tesisMCD > demandas'. It lists folders for each month from 2022-01 to 2024-04, each containing Excel files named 'rgd\_DDMMYY.xlsx'. The right panel shows the application's navigation menu with options: 'Ingesta de Datos' (selected), 'Entrenamiento', 'Evaluación', and 'Proyección Futura'. The main content area is titled '1. Ingesta y Gestión de Datos' and includes instructions on the required file format (Excel, rgd\_DDMMYY.xlsx) and data placement (A12:A36 for hourly data, L12:L36 for demand). A 'Cargar Archivos Excel' section provides a drag-and-drop area and a 'Browse files' button. Below this, three files are listed: 'rgd\_3112...', 'rgd\_3012...', and 'rgd\_2912...', each with a progress bar and a close button. The 'Estado Actual de la Base de Datos' section shows a 'Zona de Peligro: Eliminar Datos' button and displays 'Registros Diarios Totales' as 0 and 'Última Fecha'.

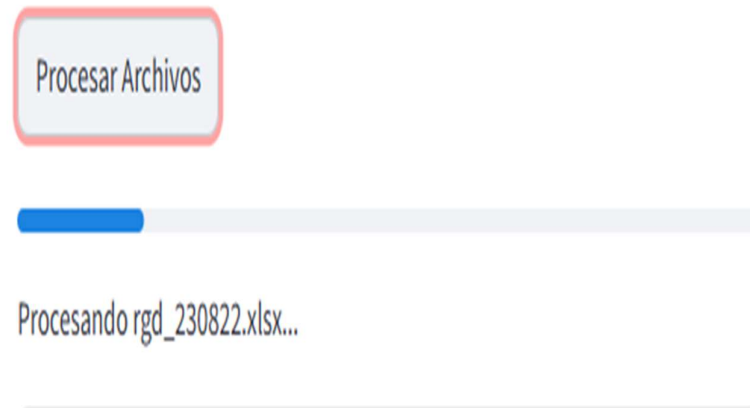
*Nota.* Elaboración propia.

Las carpetas que se pueden ver en la gráfica anterior contienen los archivos Excel mostrados anteriormente en la figura 6 con el patrón en su nombre “rgd\_DDMMYY.xlsx”, una vez culminada la carga se habilita el botón de procesar datos, el sistema procesará, limpiará y almacenará los datos en una base de datos local (`DuckDB`).



**Figura 18**

*Procesamiento de archivos*



*Nota.* Elaboración propia.

El resultado del procesamiento y consolidación de los archivos procesados, el sistema muestra una estadística y una gráfica que refleja el trabajo de limpieza e imputación de los datos.

**Figura 19***Resumen de carga de datos*

## Estadísticas de Calidad de Datos

Filas Totales (Horarias)

32925

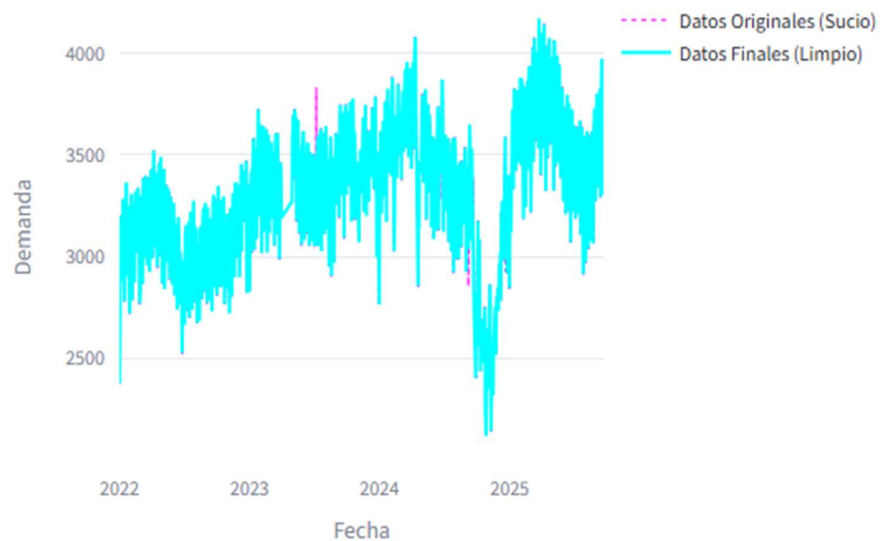
Outliers Detectados

12

Nulos Imputados

12

### Calidad de Datos: Comparativa Original vs Limpio

*Nota.* Elaboración propia.

En esta vista preliminar se puede apreciar como la limpieza de datos hace su trabajo, es decir quita los outliers e imputa datos faltantes en este grupo de datos alimentado al sistema, una vez revisado pre-procesamiento podemos proseguir con el almacenamiento de datos en la opción “guardar en base de datos” para que el data set quede disponible para el posterior tratamiento.

**Figura 20**

*Almacenamiento de dataset en base de datos*



*Nota.* Elaboración propia.

Como se puede observar en la figura anterior el sistema de igual manera muestra el resumen de la carga de datos y la respectiva gráfica de la data que se guardó en la base de datos.

### 4.3 Entrenamiento

Para la etapa de entrenamiento lo primero que se debe tomar en cuenta es la configuración del horizonte de predicción que esta expresado en días hacia el futuro, se debe seleccionar la fecha de corte para separar los conjuntos de entrenamiento y prueba.

**Figura 21**

*Panel de configuración para entrenamiento*

**UIDE ElectricForecast**  
Powered by Arizona State University®

## 2. Entrenamiento y Configuración

### Configuración de Experimento

Parámetros Generales

Horizonte de Predicción (Días)

180

Fecha de Corte (Train/Test)

2024/12/31

Hiperparámetros Modelos

☒ Prophet: Incluir Feriados (EC)

---

XGBoost

☒ Usar Optimización Bayesiana (Auto-Tune) ?

Se buscarán los mejores parámetros automáticamente.

Entrenar Modelos

☐ Entrenando Prophet y XGBoost...

*Nota.* Elaboración propia.

Una vez seleccionado el horizonte de predicción y la fecha de corte, también realizamos la configuración de los modelos que vamos a entrenar Prophet o XGBoost.

Con estos pasos realizados proseguimos con el entrenamiento al dar click en el botón de “Entrenar modelos”.

Saltará un mensaje que sugiere ir a la siguiente etapa para la revisión de resultados.

#### 4.4 Evaluación

En esta sección nosotros podemos comparar los resultados de los modelos entrenados contra la data de prueba, adicional muestra una tabla comparativa con las métricas de evaluación RMSE, MAE y MAPE, con lo cual vamos a decidir cuál es el mejor modelo para nuestro problema.

**Figura 22**

*Panel de evaluación de resultados*



### 3. Evaluación de Resultados (Backtest)

#### Tabla de Métricas

	Modelo	RMSE	MAE	MAPE
0	Prophet	2,505.1683	2,328.1172	0.6376
1	XGBoost	228.3674	180.2181	0.0484

 **Modelo más preciso (menor RMSE): XGBoost**

*Nota.* Elaboración propia.

Como se puede apreciar en el modelo anterior adicional a la tabla que evidencia los resultados, el sistema también realiza la evaluación y automáticamente muestra cuál de los 2 modelos tiene un mejor rendimiento.

De igual manera nuestro sistema muestra de forma gráfica la diferencia de los modelos vs la data real.

### Figura 23

*Panel de evaluación de resultados*

#### Backtesting: Comparativa de Modelos



*Nota.* Elaboración propia.

En la gráfica anterior se muestra claramente cuál es el modelo que más se acerca a la tendencia histórica real, por ende, indica el mejor rendimiento.

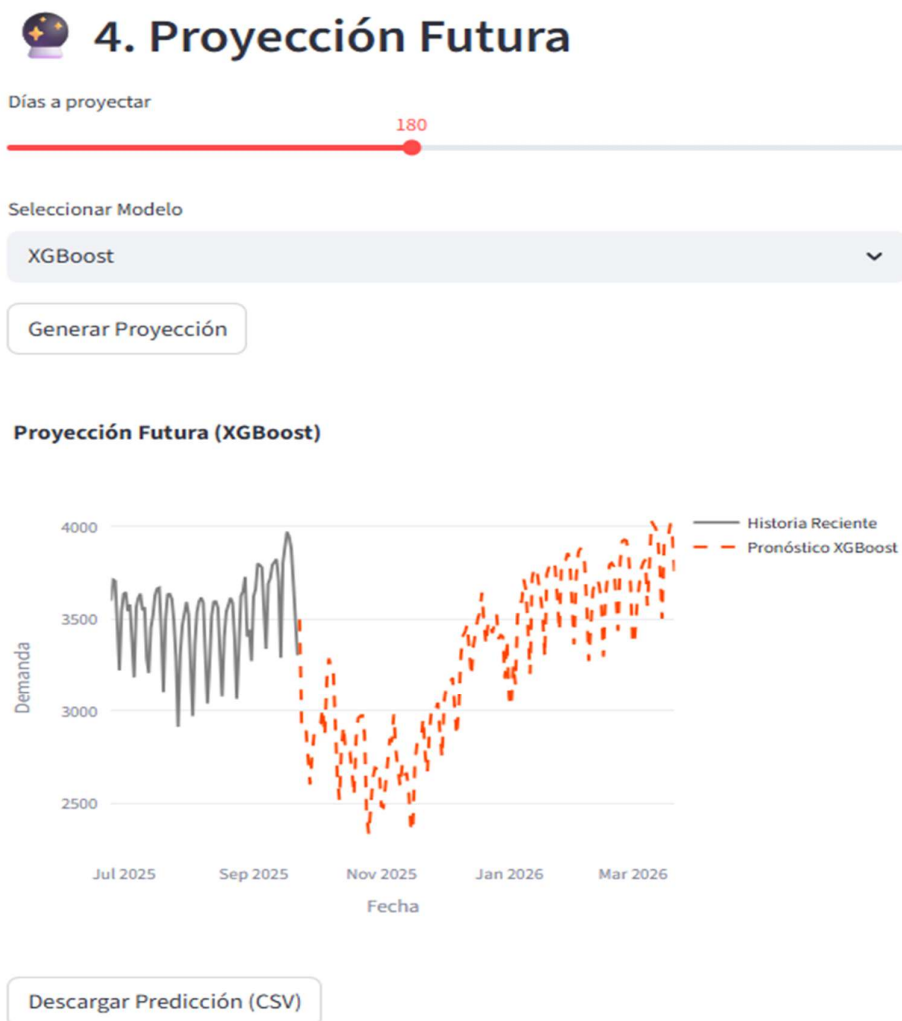
### 4.5 Proyección futura

En esta última sección se puede realizar la proyección futura de la demanda energética en la cual podemos seleccionar desde 30 días hasta 365 días, para ello vamos a seleccionar el

modelo ganador en la sección anterior que es XGBoost y finalmente generamos la proyección.

**Figura 24**

*Proyección de consumo a 180 días*



*Nota.* Elaboración propia.

Para efectos de análisis minucioso se puede exportar la proyección en un archivo CSV con el botón designado para el efecto.

#### 4.6 Análisis de resultados

Una vez realizado el modelado y del análisis de la tabla 4, sus métricas evidencian una marcada diferencia en los rendimientos de los dos modelos evaluados lo cual nos indica que

XGBoost supera por mucho en rendimiento de Prophet en este caso específico de predicción de consumo eléctrico.

Por lo tanto, vemos que el modelo XGBoost es más potente para interpretar relaciones no lineales y los patrones determinados por el consumo eléctrico.

**Tabla 4**

*Métricas de evaluación de modelos*

MODELO	RMSE	MAE	MAPE
PROPHET	2 505.1682	2 328.1172	0.6376
XGBOOST	228.3674	180.2181	0.0484

*Nota.* Elaboración propia.

Las métricas presentadas permiten evaluar de manera objetiva el desempeño de los modelos predictivos implementados, considerando distintos tipos de error. Cada métrica aporta una perspectiva complementaria sobre la calidad de las predicciones, lo que resulta fundamental para una comparación justa entre modelos con enfoques diferentes.

Analizando las métricas de rendimiento vemos que:

**RMSE:** error que penaliza fuerte en los errores grandes, en nuestro caso debido a picos mal pronosticados, este valor es demasiado alto para el modelo **prophet** lo que da a entender que hay mucha desviación al momento de alta demanda. Por otro lado, **xgboost** al tener un valor pequeño indica que es más estable y se comporta con valores cercanos a los reales.

**MAE:** en este caso esta métrica nos está indicando un error numérico en la predicción, este valor muestra una desviación mayor con el modelo **prophet** mientras que **xgboost** es un valor menor, por ende, el menor error significa una respuesta más cercana a la realidad.



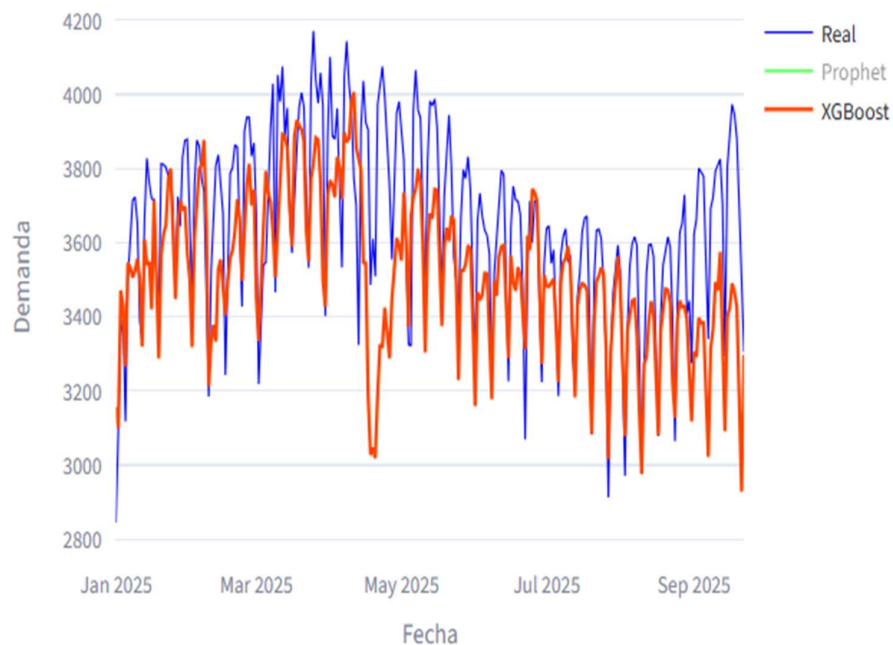
**MAPE:** Muestra el error porcentual absoluto, de esta manera entendemos directamente el menor error como un modelo más exacto, para nuestro modelado tenemos que el error de **xgboost** = 4.84% es el amplio ganador contra el error de **prophet** = 63.7%.

El marcado contraste entre los valores de error obtenidos por ambos modelos evidencia la capacidad superior de XGBoost para capturar relaciones no lineales presentes en los datos de consumo eléctrico. Este comportamiento es consistente con estudios previos que destacan el desempeño de los modelos basados en árboles de decisión ensamblados frente a métodos estadísticos tradicionales cuando se trabaja con series temporales complejas.

No obstante, es importante señalar que Prophet mantiene ventajas en términos de interpretabilidad y facilidad de ajuste, lo que lo convierte en una herramienta útil para análisis exploratorios o escenarios donde la transparencia del modelo es prioritaria.

### Figura 25

*Modelo ganador*



*Nota.* Elaboración propia.

La visualización comparativa entre los valores reales y las predicciones generadas por los modelos refuerza los resultados obtenidos a partir de las métricas numéricas. En el gráfico se observa cómo XGBoost logra un mejor ajuste a las fluctuaciones del consumo real, reduciendo la amplitud de las desviaciones en comparación con Prophet. Esta capacidad de adaptación es especialmente relevante en periodos de alta variabilidad, donde una predicción inexacta podría derivar en decisiones subóptimas de planificación energética. En consecuencia, los resultados obtenidos respaldan la selección de XGBoost como el modelo más adecuado para el escenario específico analizado en este proyecto.

## 5. CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES

### 5.1 Conclusiones

Luego de analizar el resultado de las proyecciones de ambos modelos, se concluye que el XGBoost con optimización bayesiana es el mejor modelo para predecir la demanda eléctrica en Ecuador, ya que supo ajustar su proyección a las tendencias estacionales y festivas del país. Por el contrario, el modelo Prophet no logró replicar los picos de demanda eléctrica en Ecuador, lo que indica que puede ser útil para predecir tendencias a largo plazo, pero no para predecir picos de demanda.

La implementación de **duckdb** como base de datos local permitió dar una solución eficiente para gestionar consultas de un gran volumen de datos de forma rápida y sin sobrecarga o latencia.

Esta elección tecnológica demuestra que no siempre es necesario recurrir a infraestructuras costosas o complejas para obtener resultados satisfactorio en proyectos de analítica de datos.

La correcta aplicación del EDA permitió la detección de outliers y valores faltantes para la adecuada gestión de los datos, permitiendo que los modelos no fueran entrenados con ruido y asegurando la estabilidad de las predicciones.

Cabe destacar que esta fase, aunque a menudo es subestimada, resultó determinante para evitar sesgos que hubieran comprometido la validez del modelo.

El uso del algoritmo de gradiente **xgboost** demostró una alta precisión, el resultado obtenido del  $MAPE < 5\%$  valida la correcta aplicación de los procesos ETL y EDA. Además de los resultados cuantitativos obtenidos, este trabajo demuestra que la aplicación de técnicas de ciencia de datos en el sector energético ecuatoriano es viable incluso bajo restricciones de infraestructura tecnológica.

El desarrollo de ELFO permitió combinar técnicas clásicas de limpieza de datos con algoritmos modernos de optimización bayesiana resulta en proyecciones confiables, es posible generar información confiable para apoyar procesos de planificación energética. Este aporte resulta relevante en contextos donde el acceso a plataformas comerciales avanzadas es limitado, reforzando la idea de que soluciones basadas en software libre y arquitecturas livianas pueden generar resultados de alto impacto.

Finalmente, las conclusiones obtenidas permiten afirmar que la calidad del dato juega un rol determinante en la precisión de los modelos predictivos. Las etapas de limpieza, validación y análisis exploratorio no deben considerarse procesos secundarios, sino componentes centrales del flujo de trabajo analítico. En este sentido, el proyecto refuerza la importancia de adoptar metodologías estructuradas de análisis de datos, especialmente en aplicaciones críticas como la predicción de la demanda eléctrica, donde decisiones basadas en información imprecisa pueden tener consecuencias económicas y operativas significativas

## **5.2 Recomendaciones**

Aunque el sistema es funcional, futuros trabajos podrían explorar redes neuronales recurrentes (LSTM) para capturar dependencias a más largo plazo, o integrar variables exógenas como temperatura o PIB para refinar aún más la predicción.

Asimismo, sería valioso evaluar el comportamiento del modelo ante escenarios atípicos, como los vividos durante la pandemia de COVID-19 para fortalecer su capacidad de respuesta frente a eventos imprevistos.

Como recomendación técnica, se sugiere ampliar el horizonte temporal del análisis incorporando datos históricos de mayor longitud, lo que permitiría capturar ciclos de consumo más extensos y mejorar la estabilidad de las predicciones. Un mayor volumen de datos también facilitaría el entrenamiento de modelos más complejos, como redes neuronales profundas, que requieren grandes cantidades de información para alcanzar un desempeño

óptimo. Esta ampliación temporal podría fortalecer la capacidad del sistema para anticipar cambios estructurales en el comportamiento del consumo eléctrico nacional.

Igualmente, desde una perspectiva institucional, se recomienda que las entidades responsables del sector energético promuevan la estandarización de los formatos de reporte de consumo eléctrico. La existencia de estructuras de datos homogéneas facilitaría la integración de información proveniente de distintas fuentes y reduciría el esfuerzo asociado a los procesos de limpieza y transformación. Asimismo, el acceso oportuno a datos actualizados permitiría implementar sistemas de predicción en tiempo casi real, fortaleciendo los procesos de toma de decisiones estratégicas en el sector energético. Por ejemplo, se sugiere que las entidades gubernamentales vinculadas al sector energético, como el Ministerio de Energía y Recursos Naturales No Renovables, consideren la adopción de herramientas predictivas similares a las desarrolladas en este estudio. Los modelos basados en machine learning dentro de los procesos de planificación podría mejorar significativamente la capacidad de anticipación ante variaciones en el consumo, especialmente en épocas de alta demanda como las temporadas festivas o los meses de mayor actividad industrial.

Para lograr esto, se sugiere establecer alianzas con las empresas distribuidoras que faciliten el acceso a datos actualizados de manera periódica y estandarizada.

Se podrían realizar estudios complementarios para determinar el crecimiento de usuarios ya que eso afecta directamente a la demanda de energía y ayudaría a la proyección de incremento en la generación eléctrica o compra de energía a países vecinos.

Finalmente, se insta a la posibilidad de incorporar este tipo de herramientas predictivas como apoyo permanente en la planificación energética nacional. La utilización de modelos de predicción de demanda no debe limitarse a ejercicios académicos, sino que puede integrarse como un componente clave en la gestión operativa y estratégica del sistema eléctrico. En este contexto, ELFO podría servir como base para el desarrollo de soluciones

más amplias orientadas a la optimización del uso de recursos energéticos y a la mitigación de riesgos asociados a escenarios de alta incertidumbre.

## REFERENCIAS BIBLIOGRÁFICAS

- Admin. (2020, November 30). *Análisis en big data: Métodos de minería de datos y herramientas*. DataHack. <https://www.datahack.es/analisis-big-data/>
- Adu-Poku, A., Siabi, E. K., Otchere, N. O., Effah, F. B., Awafo, E. A., Kemausuor, F., & Yazdanie, M. (2024). Impact of drought on hydropower generation in the Volta River basin and future projections under different climate and development pathways. *Energy and Climate Change*, 5, 100169.
- Almeida, M. (2024, May 22). *Pandas Python: Qué es, para qué sirve y cómo descargarlo*. Alura Cursos Online. <https://www.aluracursos.com/blog/pandas-python-que-es-para-que-sirve-como-descargar>
- AP. (2024). *Ecuador expands power cuts to 14 hours a day due to drought*. <https://apnews.com/article/ecuador-power-cuts-drought-59690c310e7c30136aad2ac7a75a043e>
- ARCERNNR. (2022). *Estadística del sector eléctrico ecuatoriano 2021*. Ministerio de Ambiente y Energía. <https://www.ambienteyenergia.gob.ec/la-arcernnr-presenta-la-estadistica-del-sector-electrico-ecuatoriano-2021/>
- ARCONEL. (2025). Panorama Eléctrico. *Revista Panorama Eléctrico*. <https://arconel.gob.ec/revista-panorama-electrico/>
- Arévalo, W. (2024). Large-scale integration of renewable energies by 2050 through demand prediction with ANFIS: *Ecuador case study*. *Energy*, 286, 129446. <https://doi.org/10.1016/j.energy.2023.129446>
- Aurora. (2024a, June 28). *Matplotlib: La herramienta esencial para visualización en data science*. ID Digital School. <https://iddigitalschool.com/bootcamps/matplotlib-la-herramienta-esencial-para-visualizacion-en-data-science/>

- Aurora. (2024b, December 10). *¿Qué es seaborn? La librería esencial para visualización de datos en Python*. ID Digital School. <https://iddigitalschool.com/bootcamps/que-es-seaborn-la-libreria-esencial-para-visualizacion-de-datos-en-python/>
- Berndt, E. (1991). *The practice of econometrics: classic and contemporary*. Addison-Wesley.
- Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). *Time series analysis: forecasting and control*. John Wiley & Sons, Inc.
- [https://books.google.com/books/about/Time\\_Series\\_Analysis.html?hl=es&id=rNt5CgAAQBAJ](https://books.google.com/books/about/Time_Series_Analysis.html?hl=es&id=rNt5CgAAQBAJ)
- Brailey, B. (2024). Transforming grid operations with accurate short-term energy predictions. DNV. <https://www.dnv.com/article/transforming-grid-operations-with-accurate-short-term-energy-predictions/>
- Candelario Gómez Cosme, F. X., & Albán Villacrés, L. E. (2024). *Aplicación de inteligencia artificial para pronosticar la demanda energética del sector de Primavera 2*.
- CELEC. (2022). *Balance Energético Nacional 2022*. [www.rekursosyenergia.gob.ec](http://www.rekursosyenergia.gob.ec)
- CENACE. (2018). *Informe Anual 2018*. <https://www.cenace.gob.ec/informe-anual-2018/>
- CENACE. (2025). *Biblioteca – Operador Nacional de Electricidad CENACE*. <https://www.cenace.gob.ec/biblioteca/>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chamorro, J., & Mera, E. (2025). Estudio de la crisis energética en el Ecuador por la dependencia en la generación de energía hidráulica. *Revista Científica INGENIAR*, 168–186. <https://journalingeniar.org/index.php/ingeniar/article/view/286>



- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, Y., Gong, W., Obrecht, C., & Kuznik, F. (2025). A review of machine learning techniques for building electrical energy consumption prediction. *Energy and AI*, 100518.
- Chicaiza, J., & Carrillo, A. (2025). *Uso de algoritmos de machine learning para predicción de consumo de energía eléctrica*. [Universidad Técnica Particular de Loja].  
<http://dspace.utpl.edu.ec/handle/29.500.19856/71492>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734.  
<https://doi.org/10.3115/v1/d14-1179>
- Choi, E., Cho, S., & Kim, D. K. (2020). Power demand forecasting using long short-term memory (LSTM) deep-learning model for monitoring energy sustainability. *Sustainability*, 12(3), 1109.
- Chong, M., & Aguilar, R. (2016). Proyección de series de tiempo para el consumo de la energía eléctrica a clientes residenciales en Ecuador. *Revista Tecnológica*, 29(1).  
<https://doaj.org/article/8bc52ceef1c34005ad0b96abe32fe1d3>
- Cole Stryker, F. L. D. B. M. S. (2025). *La guía de machine learning 2025*. Wwww.Ibm.Com/Us-En. <https://www.ibm.com/mx-es/think/machine-learning#605511093>
- Comisión Económica para América Latina y el Caribe (CEPAL). (2020). *La transición energética en América Latina y el Caribe*. Naciones Unidas.

Cristina Ortega. (2023, May 22). *Matriz de correlación: Qué es, cómo funciona y ejemplos*.

QuestionPro. <https://www.questionpro.com/blog/es/matriz-de-correlacion/>

Ember. (2025). Global Electricity Review 2025. *EMBER*. <https://ember-energy.org/latest-insights/global-electricity-review-2025/>

Equipo de contenidos de GoDaddy. (2025, March 27). *Python: Qué es, para qué sirve y por qué aprenderlo*. [www.godaddy.com/es](http://www.godaddy.com/es).

<https://www.godaddy.com/resources/latam/desarrollo/python-que-es>

Estrella, D., & Pazuña, W. (2021). *Análisis y proyección del consumo de energía eléctrica ante confinamiento por emergencia sanitaria en el área de concesión de la empresa eléctrica provincial Cotopaxi. Caso de estudio cantón Latacunga*. [Universidad Técnica de Cotopaxi]. <http://repositorio.utc.edu.ec/handle/27000/8016>

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37.  
<https://doi.org/10.1609/aimag.v17i3.1230>

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.

Fernández-Avilés, G., & M. J.-M. (2024). *Fundamentos de ciencia de datos con R*. McGraw-Hill.

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (2nd ed.). Analytics Press.

Fose, N., Singh, A. R., Krishnamurthy, S., Ratshitanga, M., & Moodley, P. (2024). Empowering distribution system operators: A review of distributed energy resource forecasting techniques. *Heliyon*, 10(15).

- Gallo, A., Pérez, F., & Salinas, D. (2021). Data Mining and Short-Term Projection of Power Demand in the Ecuadorian Electric System. *Revista Técnica Energía*, 18(1), 72–85.  
<https://doi.org/10.37116/REVISTAENERGIA.V18.N1.2021.461>
- Gellings, C., & Chamberlin, J. (1988). Demand-side management: concepts and methods. In *Demand-side management: concepts and methods*. Fairmont Press.  
[https://books.google.com/books/about/Demand\\_side\\_Management.html?hl=es&id=LYPtAAAAMAAJ](https://books.google.com/books/about/Demand_side_Management.html?hl=es&id=LYPtAAAAMAAJ)
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Grainger, J., & Stevenson, W. (1994). *Power System Analysis*. McGraw-Hill.  
[https://books.google.com/books/about/Power\\_System\\_Analysis.html?hl=es&id=a1OzQgAACAAJ](https://books.google.com/books/about/Power_System_Analysis.html?hl=es&id=a1OzQgAACAAJ)
- Guachimboza-Davalos, J. I., Llanes-Cedeño, E. A., Rubio-Aguilar, R. J., Peralta-Zurita, D. B., & Núñez-Barrionuevo, O. F. (2021). Prediction of Monthly Electricity Consumption by Cantons in Ecuador Through Neural Networks: A Case Study. *Advances in Intelligent Systems and Computing*, 1273 AISC, 28–37. [https://doi.org/10.1007/978-3-030-59194-6\\_3](https://doi.org/10.1007/978-3-030-59194-6_3)
- Guaman, A., Ramirez, J., Mayorga, B., Aviles, F., & Gallardo, C. (2019). Short-Term Load Forecasting in the Distribution System of the Electric Company of Ambato (EEASA) Based on Big Data Criteria. *Proceedings - 2019 International Conference on Information Systems and Computer Science, INCISCOS 2019*, 23–30.  
<https://doi.org/10.1109/INCISCOS49368.2019.00013>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

- Hasan, M., Mifta, Z., Papiya, S. J., Roy, P., Dey, P., Salsabil, N. A., ... & Farrok, O. (2025). A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications. *Energy Conversion and Management*: X, 100922.
- Hasanat, S. M., Ullah, K., Yousaf, H., Munir, K., Abid, S., Bokhari, S. A. S., Aziz, M. M., Naqvi, S. F. M., & Ullah, Z. (2024). Enhancing Short-Term Load Forecasting With a CNN-GRU Hybrid Model: A Comparative Analysis. *IEEE Access*, 12, 184132–184141. <https://doi.org/10.1109/ACCESS.2024.3511653>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/NECO.1997.9.8.1735>
- Hong, T., Pinson, P., & Fan, S. (2016). Global energy forecasting competition 2012. *International Journal of Forecasting*, 32(3), 896–913. <https://doi.org/10.1016/j.ijforecast.2016.02.001>
- Hossain, M. S., & Mahmood, H. (2020). Short-Term Load Forecasting Using an LSTM Neural Network. *2020 IEEE Power and Energy Conference at Illinois, PEI 2020*. <https://doi.org/10.1109/PEI48348.2020.9064654>
- Huang, Z., & Yi, Y. (2024). Short-Term Load Forecasting for Regional Smart Energy Systems Based on Two-Stage Feature Extraction and Hybrid Inverted Transformer. *Sustainability* 2024, Vol. 16, Page 7613, 16(17), 7613. <https://doi.org/10.3390/SU16177613>

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.).

OTexts.

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

OTexts.

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

IEA. (2025). Electricity 2025. *IEA*. <https://www.iea.org/reports/electricity-2025/demand>

IIGE. (2020). *Instituto de Investigación Geológico y Energético*.

<https://www.geoenergia.gob.ec/category/noticias/page/25/>

Ingeniería. (2024, September 17). *NumPy: biblioteca imprescindible para la ciencia de datos*.

Universidadeuropea.Com. <https://universidadeuropea.com/blog/numpy/>

Instituto de Investigación Geológico y Energético (IIGE). (2020). *Balance energético nacional del Ecuador*. IIGE.

International Energy Agency (IEA). (2022). *World energy outlook 2022*. IEA.

IPCC. (2023). Energy Systems. *Climate Change 2022 - Mitigation of Climate Change*, 613–746. <https://doi.org/10.1017/9781009157926.008>

Jaramillo, M. (2024). Time Series Analysis for Electrical Demand Forecasting. *Digital Technology for Smart Grid Innovative Algorithmic Solutions for Engineering Problems*, 166–182. <https://doi.org/10.17163/ABYAUPS.44.351>

Jaramillo, M., & Llamuca, S. (2022). A Proposed Model for Electricity Demand Forecasting in Ecuador Considering Akaike Criterion. *Smart Innovation, Systems and Technologies*, 252, 345–355. [https://doi.org/10.1007/978-981-16-4126-8\\_32](https://doi.org/10.1007/978-981-16-4126-8_32)

- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.
- Kundur, P. (1994). *Power system stability and control*. McGraw-Hill.
- Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221, 386–405. <https://doi.org/10.1016/J.APENERGY.2018.02.069>
- López, J., Valero, F., & Senabre, C. (2019). Machine learning approaches for electricity demand forecasting. *Energy Systems*, 10(3), 735–760. <https://doi.org/10.1007/s12667-018-0300-6>
- Mahmood, A. M., Abdul Zahra, M. M., Hamed, W., Bashar, B. S., Abdulaal, A. H., Alawsi, T., & Adhab, A. H. (2022). Electricity Demand Prediction by a Transformer-Based Model. *Majlesi Journal of Electrical Engineering*, 16(4), 97–102. <https://doi.org/10.30486/MJEE.2022.696520>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Martins, C., & Macêdo, B. (2024). Electricity energy demand prediction using computational intelligence techniques. *Latin-American Journal of Computing*, 11(2), 80–88. <https://lajc.epn.edu.ec/index.php/LAJC/article/view/392>
- Michael Goodwin. (n.d.). ¿Qué es una API? *www.ibm.com/us-en*. Retrieved December 14, 2025, from <https://www.ibm.com/mx-es/think/topics/api>
- Ministerio de Energía y Recursos Naturales No Renovables. (2021). *Plan maestro de electricidad del Ecuador*. Gobierno del Ecuador.

Montero Laurencio, R., Torres Breff, O., Marrero Ramírez, S., & Jiménez Jiménez, D. (2024).

Predicción de Consumo y Demanda de Electricidad Mediante Redes Neuronales

Artificiales y Algoritmo Iterativo. *Revista Poli técnica*, 54(3), 45–58.

<https://doi.org/10.33333/RP.VOL54N3.05>

Mora, A., & Toaquiza, S. (2021). *Predicción de demanda eléctrica utilizando redes neuronales*

*artificiales para CNEL EP – unidad de negocios agencia Quevedo* [Universidad Técnica

Estatad de Quevedo]. <https://repositorio.uteq.edu.ec/handle/43000/6595>

Muñoz, A., Sánchez-Úbeda, E. F., Cruz, A., & Marín, J. (2010). *Short-term Forecasting in*

*Power Systems: A Guided Tour*. 129–160. [https://doi.org/10.1007/978-3-642-12686-4\\_5](https://doi.org/10.1007/978-3-642-12686-4_5)

Oleh Dubetcky. (2024, July 15). *Project Management for Data Science: KDD, SEMMA and*

*CRISP-DM*. Medium.com. [https://oleg-dubetcky.medium.com/project-management-for-](https://oleg-dubetcky.medium.com/project-management-for-data-science-kdd-semma-and-crisp-dm-fe9d03d3ab6c)

[data-science-kdd-semma-and-crisp-dm-fe9d03d3ab6c](https://oleg-dubetcky.medium.com/project-management-for-data-science-kdd-semma-and-crisp-dm-fe9d03d3ab6c)

Ortiz, E., & García, M. (2015). *Pronóstico de la demanda eléctrica residencial basado en el*

*modelo de regresión adaptativa multivariante spline (MARS)*.

<http://dspace.ups.edu.ec/handle/123456789/11290>

Pansini, A. (2020). *Electrical distribution engineering*. River Publishers.

[https://books.google.com/books/about/Electrical\\_Distribution\\_Engineering\\_Thir.html?hl](https://books.google.com/books/about/Electrical_Distribution_Engineering_Thir.html?hl)

[=es&id=5SAJEAAAQBAJ](https://books.google.com/books/about/Electrical_Distribution_Engineering_Thir.html?hl=es&id=5SAJEAAAQBAJ)

Pérez-López, D., Muqtadir, A., Li, B., Qi, B., Ge, L., Du, N., & Lin, C. (2025). Demand

Response Potential Forecasting: A Systematic Review of Methods, Challenges, and

Future Directions. *Energies* 2025, Vol. 18, Page 5217, 18(19), 5217.

<https://doi.org/10.3390/EN18195217>

Pressman, R. S., & Maxim, B. R. (2020). *Software engineering: A practitioner's approach* (9th

ed.). McGraw-Hill Education.

Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

Radicelli, C., & Parra, V. (2025). *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.*

<http://dspace.unach.edu.ec/handle/51000/15221>

Richardson, I., Thomson, M., Infield, D., & Clifford, C. (2010). Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 42(10), 1878–1887.

<https://doi.org/10.1016/J.ENBUILD.2010.05.023>

Salazar, G., & Panchi, B. (2014). Análisis de la Evolución de la Demanda Eléctrica en el Ecuador Considerando el Ingreso de Proyectos de Eficiencia Energética. *Revista Politécnica*, 33(1).

[https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista\\_politecnica2/article/view/218](https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/218)

Sánchez, E., & Sinche, J. (2024). *Predicción de demanda eléctrica mediante Redes Neuronales para la anticipación de las necesidades de suministro.*

<http://www.dspace.espol.edu.ec/handle/123456789/62294>

Serrano-Guerrero, X., Briceño-León, M., Clairand, J. M., & Escrivá-Escrivá, G. (2021). A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. *Applied Energy*, 297, 117173.

<https://doi.org/10.1016/J.APENERGY.2021.117173>

Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for electricity demand using machine learning techniques. *Energy Reports*, 6, 325–335.

<https://doi.org/10.1016/j.egy.2020.01.006>



- Sharda, R. D. (2018). *Inteligencia de big data, análisis y ciencia de datos: una perspectiva gerencial*. Pearson Educación.
- Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, & artificial intelligence: Systems for decision support* (11th ed.). Pearson.
- Silva, S., Soares, I., & Pinho, C. (2020). Climate change impacts on electricity demand. *Utilities Policy*, 67, 101115. <https://doi.org/10.1016/j.jup.2020.101115>
- Song, K. M., Kim, T. G., Cho, S. M., Song, K. Bin, & Yoon, S. G. (2025). XGBoost-Based Very Short-Term Load Forecasting Using Day-Ahead Load Forecasting Results. *Electronics* 2025, Vol. 14, Page 3747, 14(18), 3747.  
<https://doi.org/10.3390/ELECTRONICS14183747>
- Stott, M. (2024, noviembre 11). Latin America's hydro power bet suffers effects of climate change. *Financial Times*. <https://www.ft.com/content/059a0657-d508-4c34-a517-57e5cca0ee4c>
- Suntaxi, J., Salazar, Y., & Loor, R. (2019). Pronóstico de la demanda en energía y potencia del Sistema Eléctrico Quito. *Revista Técnica Energía*, 15(2), 12–21.  
<https://doi.org/10.37116/REVISTAENERGIA.V15.N2.2019.371>
- Tigua, Y., Cadena-Barrera, J., Castro-Verdezoto, P., & Ortega-Pacheco, D. (2025). Energy forecast of Ecuador's Tertiary Sector to 2040 Using the LEAP Model. *EASI: Engineering and Applied Sciences in Industry*, 4(2), 1–8.  
<https://doi.org/10.53591/EASI.V4I2.2596>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*. <https://arxiv.org/abs/1706.03762>

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probability and statistics for engineers and scientists* (9th ed.). Pearson.

Weron, R. (2014). *Electricity price forecasting*. *International Journal of Forecasting*, 30(4), 1030–1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art. *International Journal of Forecasting*, 30(4), 1030–1081.  
<https://doi.org/10.1016/j.ijforecast.2014.08.008>

Winters, P. R. (1960). *Forecasting sales by exponentially weighted moving averages*. *Management Science*.

World Bank. (2021). *Tracking SDG 7: The energy progress report*. World Bank.

Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)

**ANEXOS**

<https://github.com/jlvd/ElectricForecast/tree/master>