

Implementación de un modelo basado en datos para la predicción del consumo energético en una vivienda de ingresos medios-altos

Jorge Enrique Sierra Navarro y Elibardo Valencia Saavedra

Trabajo de grado para optar al título de Ingeniero Electricista

Director

Jorge Luis Cardenas Rangel

PhD (c) en Ingeniería, Área Ingeniería Eléctrica

Co-Director

German Alfonso Osma Pinto

Doctor en Ingeniería, Área Ingeniería Eléctrica

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T)

Ingeniería Eléctrica

Bucaramanga

2026

Dedicatoria

Dedico este trabajo de grado a mi familia, por su apoyo constante, su paciencia y su confianza a lo largo de todo este proceso. Su acompañamiento ha sido fundamental para poder culminar esta etapa de formación. Asimismo, dedico este trabajo a todas las personas que, de una u otra forma, creyeron en este proyecto y me motivaron a seguir adelante incluso en los momentos de mayor dificultad.

Elibardo Valencia Saavedra

Quisiera dedicar este trabajo en primer lugar a mis padres, Alexandra y Javier, por educarme y su apoyo incondicional durante toda mi vida, a mi hermana Valentina por ser la mejor amiga que me dio la vida. A Ana por estar presente cuando más lo he necesitado en los últimos años con su amor y paciencia, y finalmente al director de este trabajo Jorge por su orientación, tiempo y enseñanza durante este trabajo.

Jorge Enrique Sierra Navarro

Agradecimientos

Agradecemos en primer lugar a Dios por darnos la fortaleza, la disciplina y la perseverancia necesarias para completar este trabajo de grado.

Expresamos un especial agradecimiento a nuestro director de trabajo de grado Jorge Luis Cardenas Rangel, por su orientación, su acompañamiento académico y sus valiosos aportes durante el desarrollo del proyecto. Sus observaciones y sugerencias fueron fundamentales para mejorar la calidad del trabajo y orientar correctamente el enfoque de la investigación.

También agradecemos a la Universidad Industrial de Santander por brindarnos la formación académica y las herramientas necesarias para llevar a cabo este proyecto, así como a los docentes que contribuyeron a nuestra formación profesional.

De manera particular, agradecemos a las personas que facilitaron la información, los datos y el acceso a la vivienda utilizada como caso de estudio, sin los cuales no habría sido posible realizar el análisis desarrollado en este trabajo.

Finalmente, agradecemos a nuestros compañeros, amigos y familiares por su apoyo constante, su comprensión y su ánimo durante todo el proceso de elaboración de este trabajo de grado.

Tabla de contenido

Introducción 10

1. Problema..... 11

2. Objetivos 12

2.1.1 Objetivo general..... 12

2.1.2 Objetivos específicos..... 12

3. Descripción del documento 13

4. Modelado energético 14

4.1 Descripción de la vivienda 14

4.2 Creación del modelo 16

4.3 Integración de datos en el modelo 17

4.4 Ajuste del modelo 22

5. Modelo basado en datos 25

5.2 Revisión y descripción de los modelos de predicción..... 27

5.3 Elaboración del modelo..... 31

5.3.1 Implementación del modelo de Regresión Lineal..... 32

5.3.3 Implementación del modelo de MLP 34

5.3.4 Implementación del modelo de XGBoost 35

5.4 Evaluación de rendimiento de los modelos 36

6. Comparación de modelos de predicción..... 40

6.1 Comparación de los modelos..... 40

6.2 Error en los modelos 43

7. Conclusiones..... 45

Referencias Bibliográficas 48

Lista de tablas

Tabla 1. Inventario de equipos de la vivienda	15
Tabla 2. Simulación 1 en DesignBuilder	23
Tabla 3. Simulación 4 en DesignBuilder	24
Tabla 4. Resultados de demanda computacional de los modelos basados en datos	36
Tabla 5. Resultados de los modelos basados en datos	37
Tabla 6. Métricas de comparación de DesignBuilder.....	42

Lista de figuras

Figura 1. Edificio Qatar 15

Figura 2. Plano planta del apartamento..... 15

Figura 3. Plano de planta de piso del Edificio Qatar y construcción en DesignBuilder 16

Figura 4. Modelo geométrico Edificio Qatar 17

Figura 5. Muro de fachada 18

Figura 6. Particiones internas generales..... 18

Figura 7. División interna distinta..... 19

Figura 8. Puertas intenas del apartamento del Edificio Qatar..... 20

Figura 9. Vista renderizada del Edificio Qatar 21

Figura 10. Participación por uso final del apartamento 25

Figura 11. Ubicación de sensores y equipo de medida de energía..... 27

Figura 12. Implementación del modelo de Regresión Lineal..... 32

Figura 13. Implementación del modelo de Random Forest 33

Figura 14. Implementación del modelo de MLP 34

Figura 15. Implementación del modelo de XGBoost..... 35

Figura 16. Resultados de los modelos basados en datos..... 38

Figura 17. Curvas de consumo Real, DesignBuilder 42

Figura 18. Curvas de consumo Real, Random Forest y DesignBuilder 43

Figura 19. Porcentaje de error entre el modelo de Random Forest, DesignBuilder..... 45

Lista de apéndices

Los apéndices están disponibles en el Repositorio Institucional

Apéndice A. Programación de equipos

Apéndice B. Variables dataset

Apéndice C. Dataset

Apéndice D. Resultados día predicho y métricas

Apéndice E. Edificio Qatar Apartamento

Apéndice F. Código machine learning y comparación

Apéndice G. Comparación mes a mes DB

Apéndice H. Imágenes de los equipos utilizados

Resumen

Título: Implementación de un modelo basado en datos para la predicción del consumo energético en una vivienda de ingresos medios-altos*

Autores: Jorge Enrique Sierra Navarro, Elibardo Valencia Saavedra**

Palabras clave: Consumo de energía, Edificación residencial, Machine Learning, Simulación energética

Descripción: Este trabajo de grado desarrolla un estudio comparativo sobre la predicción del consumo energético de un apartamento residencial de ingresos medios-altos. Para ello, se analizaron dos enfoques de predicción: un modelo de simulación energética basado en el software DesignBuilder y modelos basados en datos (machine learning), con el objetivo de evaluar su capacidad para reproducir el comportamiento real del consumo eléctrico.

Se implementaron cuatro modelos de predicción basados en datos: Regresión Lineal, *Random Forest*, Redes Neuronales Multicapa (MLP) y XGBoost. Todos los modelos fueron entrenados y evaluados bajo una misma metodología, empleando métricas ampliamente aceptadas en la literatura, tales como: el error cuadrático medio (RMSE), el error absoluto medio (MAE), el coeficiente de determinación (R^2) y el error porcentual absoluto medio simétrico (SMAPE). A partir del análisis de estas métricas y de las curvas de consumo predichas, se determinó que, bajo las condiciones del estudio, los modelos basados en árboles de decisión presentan el mejor desempeño, destacándose *Random Forest* por su mayor capacidad para capturar de manera consistente la tendencia y la variabilidad del consumo energético.

Finalmente, se realizó una comparación entre el mejor modelo de *machine learning* y el modelo de simulación obtenido mediante DesignBuilder, el cual fue previamente calibrado de acuerdo con la ASHRAE Guideline 14. Si bien el modelo de simulación presentó valores más bajos de SMAPE, su desempeño en el resto de las métricas fue inferior al del modelo basado en datos, evidenciando limitaciones para reproducir la dinámica temporal y los picos de consumo.

* Trabajo de grado

** Facultad de Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Ingeniería Eléctrica.
Director: Jorge Luis Cardenas Rangel. Codirector: German Alfonso Osma Pinto

Abstract

Title: Implementation of a Data-Driven Model for Energy Consumption Prediction in a Middle–High Income Residential Dwelling*

Authors: Jorge Enrique Sierra Navarro, Elibardo Valencia Saavedra**

Keywords: Energy consumption, Residential building, Machine learning, Energy simulation.

Description: This bachelor's thesis presents a comparative study on the prediction of energy consumption in a middle–high income residential apartment. Two different forecasting approaches are analyzed: a physics-based energy simulation model developed using DesignBuilder, and data-driven models based on machine learning, with the aim of evaluating their ability to reproduce the actual electricity consumption behavior.

Four data-driven prediction models were implemented: Multiple Linear Regression, Random Forest, Multilayer Perceptron (MLP), and XGBoost. All models were trained and evaluated under the same methodological framework, using performance metrics widely reported in the literature, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), and Symmetric Mean Absolute Percentage Error (SMAPE). Based on the analysis of these metrics and the predicted consumption curves, the results indicate that tree-based models exhibit the best performance under the conditions of this study, with Random Forest standing out for its ability to consistently capture both the trend and variability of energy consumption.

Finally, a comparison was conducted between the best-performing machine learning model and the energy simulation model obtained using DesignBuilder, which was previously calibrated according to the ASHRAE Guideline 14. Although the simulation model achieved lower SMAPE values, its performance in the remaining metrics was inferior to that of the data-driven model, revealing limitations in reproducing the temporal dynamics and consumption peaks.

* Bachelor Thesis

** Facultad de Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Ingeniería Eléctrica. Director: Jorge Luis Cardenas Rangel. Codirector: German Alfonso Osma Pinto

Introducción

Desde el inicio de la Revolución Industrial, el consumo de energía ha experimentado un crecimiento sostenido, impulsado por el avance tecnológico, la urbanización y el aumento poblacional. A lo largo del siglo XX, el acceso a la electricidad se expandió de manera significativa, consolidándose como un elemento fundamental para el desarrollo económico y social. Esta tendencia se ha intensificado en el siglo XXI, donde la demanda energética global continúa en ascenso debido a factores como la digitalización de las actividades cotidianas, el incremento en el uso de sistemas de climatización y la construcción de edificaciones más grandes y complejas (Yesilyurt et al., 2024).

En este contexto, el sector de las edificaciones constituye uno de los principales consumidores de energía a nivel mundial, representando aproximadamente un tercio del consumo energético total (Ortega-Díaz et al., 2023; Revista UIS Ingenierías, s. f.). Dentro de este sector, el ámbito residencial destaca por su elevada demanda, siendo responsable de cerca del 30 % del consumo eléctrico total en Europa (Iqbal et al., 2018) y del 31 % de la energía final global en 2019, según el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) (Cabeza et al., 2022). Particularmente, las viviendas de ingresos medios-altos concentran una demanda creciente atribuida al uso intensivo de electrodomésticos, sistemas de climatización y tecnologías orientadas al confort (Kowshik et al., 2024).

Esta tendencia se ha visto acentuada por el crecimiento poblacional, la urbanización acelerada, el cambio climático y una mayor permanencia de los usuarios en el hogar, generando una presión adicional sobre los sistemas energéticos existentes (Kowshik et al., 2024). En el caso de Colombia, este desafío adquiere mayor relevancia, dado que el sector de edificaciones concentra una proporción significativa del consumo eléctrico del Sistema Interconectado Nacional (SIN), y se proyecta que la demanda de energía continúe creciendo de manera sostenida (Revista UIS Ingenierías, s. f.). En este escenario, resulta imprescindible desarrollar estimaciones de consumo energético cada vez más precisas, que permitan comprender el comportamiento real de la demanda y apoyar la toma de decisiones en materia de planeación, eficiencia energética y operación.

Asimismo, los modelos de predicción deben equilibrar tres factores clave: precisión, velocidad de cómputo y disponibilidad de datos, adaptándose a contextos donde no siempre se dispone de información completa o detallada. No obstante, persisten desafíos importantes, entre ellos la necesidad de mejorar la interpretabilidad de los resultados, facilitar su validación práctica

y fomentar una mayor comparación con herramientas de simulación energética consolidadas, como DesignBuilder (Ortega-Díaz et al., 2023; Zahedi et al., 2023).

Considerando este panorama, el presente trabajo propone abordar la problemática mediante la implementación de un modelo predictivo basado en datos para estimar el consumo energético en una vivienda de ingresos medios-altos, segmento caracterizado por un consumo creciente asociado al uso intensivo de sistemas de climatización, electrodomésticos y tecnologías orientadas al confort.

1. Problema

Las edificaciones presentan un aumento sostenido en la demanda energética, impulsado por factores como el cambio climático, la urbanización acelerada y la permanencia prolongada de los usuarios en sus hogares. Esta tendencia plantea desafíos significativos para el cumplimiento de metas de sostenibilidad y para el dimensionamiento adecuado de la infraestructura eléctrica, especialmente en países como Colombia, donde el sector de edificaciones representa una proporción significativa del consumo eléctrico del Sistema Interconectado Nacional (SIN) (Revista UIS Ingenierías, s. f.).

Ante este panorama, la implementación de estrategias de eficiencia energética en el entorno residencial se convierte en una prioridad. Para ello, contar con herramientas precisas para la estimación del consumo energético constituye un paso fundamental. Sin embargo, esta tarea representa un desafío técnico debido a la multiplicidad de variables involucradas y a los conocimientos especializados que típicamente requieren las herramientas convencionales, lo que se traduce en mayor tiempo y complejidad de uso (Yesilyurt et al., 2024; Revista UIS Ingenierías, s. f.; Zahedi et al., 2023).

Actualmente, este reto puede abordarse mediante dos enfoques principales. Por un lado, los modelos de ingeniería basados en principios físicos ofrecen alta capacidad de representación del comportamiento energético; no obstante, requieren entradas detalladas, información geométrica y constructiva precisa, así como procesos de modelado extensos, lo que implica mayores costos de tiempo y recursos. Además, pueden presentar limitaciones al capturar la variabilidad real del comportamiento de los usuarios (Zahedi et al., 2023; Ortega-Díaz et al., 2023).

Por otro lado, los modelos basados en datos, que emplean técnicas de aprendizaje automático (Machine Learning), se han consolidado como una alternativa prometedora. Estos modelos permiten aprender patrones directamente a partir de datos históricos de consumo y

variables asociadas —como uso de electrodomésticos, ocupación y condiciones ambientales—, logrando predicciones eficientes con menor carga de modelado físico explícito (Ortega-Díaz et al., 2023; Kowshik et al., 2024; Chen et al., 2024; Amasyali & El-Gohary, 2018).

En este contexto, resulta pertinente evaluar la calidad de los modelos basados en datos en comparación con los modelos convencionales de simulación, con el fin de determinar si es posible obtener predicciones de calidad a partir de menores requerimientos de información, tiempo de implementación y nivel de experticia del usuario.

El presente proyecto tiene como objetivo analizar y comparar dos enfoques para la predicción del consumo energético en una vivienda urbana de ingresos medios-altos ubicada en clima cálido colombiano. El primer enfoque se basa en un modelo de ingeniería desarrollado en DesignBuilder, el cual simula el comportamiento energético a partir de variables físicas y climáticas. El segundo enfoque emplea técnicas de aprendizaje automático utilizando datos provenientes de un experimento de monitoreo energético y ambiental. Para este último, se entrenaron cuatro algoritmos: Regresión Lineal (RL), Random Forest (RF), Perceptrón Multicapa (MLP) y XGBoost.

Los resultados fueron comparados considerando criterios como el error de predicción, la demanda computacional, la facilidad de implementación y el tipo de información requerida. Para la evaluación del desempeño se emplearon las métricas coeficiente de determinación (R^2), error absoluto medio (MAE), error porcentual absoluto medio simétrico (SMAPE) y raíz del error cuadrático medio (RMSE). Esta comparación permitió analizar la brecha entre modelos físicos y modelos basados en datos, aportando elementos relevantes para el diseño y la gestión energética en el sector residencial colombiano.

2. Objetivos

2.1 Objetivo general

Desarrollar un modelo basado en datos para la predicción del consumo energético en una vivienda de ingresos medios-altos y evaluar su desempeño frente a un modelo de simulación en DesignBuilder.

2.2 Objetivos específicos

1. Desarrollar un modelo energético en DesignBuilder para una vivienda de ingresos medios-altos, con el fin de predecir su consumo energético.
2. Establecer el modelo de predicción de energía basado en datos de una vivienda de ingresos medios-altos.

3. Comparar el desempeño de un modelo de predicción de energía basado en datos con respecto a su homólogo de ingeniería.

3. Descripción del documento

Este documento se estructura en cuatro capítulos principales, organizados de manera que permiten desarrollar progresivamente el análisis del consumo energético en una vivienda de ingresos medios-altos y la comparación entre enfoques de modelado energético.

La primera sección presenta la introducción del trabajo, donde se contextualiza el problema del aumento del consumo energético en el sector residencial, se justifica la relevancia del estudio y se plantea el problema de investigación. Asimismo, se definen el objetivo general y los objetivos específicos que orientan el desarrollo del proyecto.

La segunda sección aborda el modelado energético de la vivienda mediante un enfoque de ingeniería. En este capítulo se describe detalladamente la vivienda objeto de estudio, se explica el proceso de creación del modelo geométrico en DesignBuilder y se documenta la integración de los datos constructivos, operativos y climáticos. Finalmente, se presenta el proceso de ajuste y calibración del modelo energético, comparando los resultados simulados con los consumos reales para asegurar su validez.

La tercera sección desarrolla el modelo de predicción de energía basado en datos. Se describe el *dataset* utilizado, compuesto por mediciones reales de consumo y variables ambientales y operativas, y se revisan los modelos de predicción seleccionados. Posteriormente, se detalla la implementación de los modelos de Regresión Lineal, *Random Forest*, MLP y XGBoost, así como la evaluación de su desempeño mediante métricas estadísticas que permiten analizar su capacidad predictiva.

Finalmente, en la cuarta sección presenta la comparación entre el modelo de predicción basado en datos y el modelo de ingeniería. En este capítulo se contrastan los resultados obtenidos por el modelo de mejor desempeño con los resultados generados por DesignBuilder, analizando sus diferencias en términos de precisión, capacidad de representar la variabilidad del consumo y comportamiento frente a picos de demanda. A partir de esta comparación, se discuten las principales conclusiones del estudio y se evidencia el potencial de los modelos basados en datos como herramientas complementarias para el análisis energético en el sector residencial.

4. Modelado energético

En esta sección se presenta la construcción del modelo energético del apartamento a partir de sus características físicas, constructivas y operativas. Se describe la geometría de la vivienda, los materiales, las cargas internas y los perfiles de uso que se incorporan en DesignBuilder, así como el proceso de ajuste del modelo con base en los datos reales de consumo. El objetivo es contar con una representación física del comportamiento energético de la vivienda que sirva como referencia para la comparación posterior.

4.1 Descripción de la vivienda

La vivienda objeto de estudio corresponde a un apartamento ubicado en el edificio Qatar apartamentos, un conjunto residencial multifamiliar de estrato 4 de 18 pisos o niveles, situado en la Calle 48 #23-26, barrio Antonia Santos, en el centro de Bucaramanga.

La vivienda pertenece a un segmento de ingresos medios-altos, característico del estrato 4, lo que se refleja en el nivel de dotación de electrodomésticos y equipos de confort presentes en la vivienda.

En cuanto a distribución cuenta con:

- Sala común: Sala, comedor
- Cocina integral
- Tres habitaciones, siendo la principal con baño privado
- Dos baños, uno en la habitación principal y el otro para la sala común
- Balcón con vista exterior

De forma general, el apartamento del Edificio Qatar representa una vivienda urbana típica de clase media-alta en Bucaramanga, con un nivel de equipamiento eléctrico que lo convierte en un caso de estudio representativo para el análisis de consumos energéticos.

La Figura 1 presenta la fachada del edificio Qatar y la Figura 2 muestra el plano de planta del apartamento que es objeto de estudio.

Figura 1.

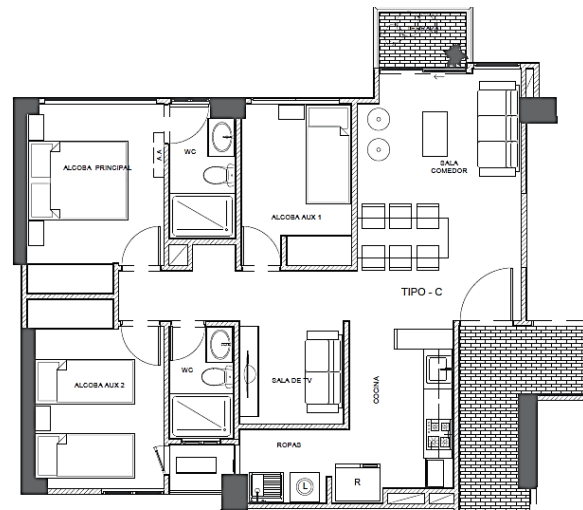
Edificio Qatar



Nota. Vista de la fachada del Edificio Qatar

Figura 2.

Plano planta del apartamento



Nota. Distribución arquitectónica del apartamento

El apartamento analizado tiene un área construida de 75 m² y se encuentra actualmente habitado por dos personas. Las cargas eléctricas se encuentran distribuidas por zonas funcionales, de acuerdo con su uso y localización dentro de la vivienda, tal como se presenta en la Tabla 1.

Tabla 1.

Inventario de equipos de la vivienda

Zona	Carga	Cantidad	Potencia [W]
Zona común (Sala, comedor, estudio)	Iluminación LED	3	36
	Computador de mesa	1	110
	TV/Entretenimiento	1	60
	PC (Portátil)	1	40
	Router	1	10
Lavandería	Iluminación LED	1	12
	Lavadora	1	400
	Medidor	1	4
Alcoba principal	Iluminación LED	1	12
	Cargadores (celular/Tablet)	1	20
	Aire acondicionado mini-split 12000 BTU	1	1800
Alcoba auxiliar 1	Iluminación LED	1	12
Alcoba auxiliar 2	Iluminación LED	1	12

Baño alcoba principal	Iluminación LED	1	12
Baño auxiliar	Iluminación LED	1	12

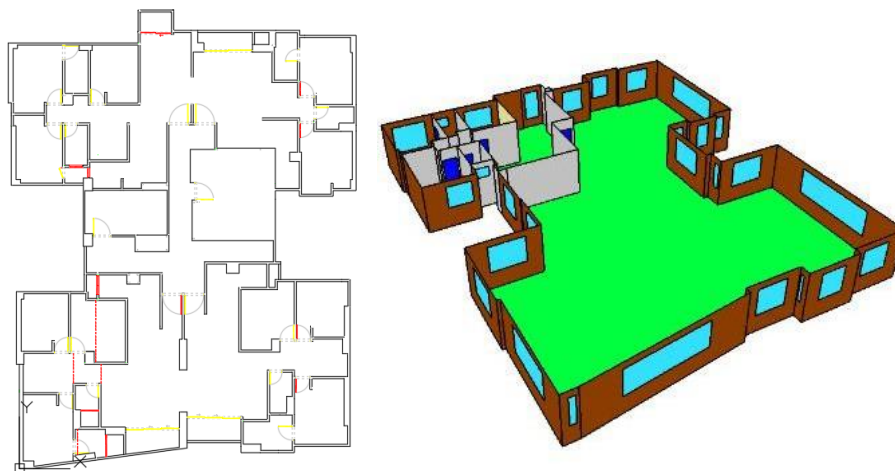
Nota. Equipos eléctricos considerados en el modelo energético del apartamento.

4.2 Creación del modelo

El modelo energético fue desarrollado en el software DesignBuilder versión 6, a partir de los planos arquitectónicos del apartamento y del piso completo, suministrados por la empresa constructora Oteco. No obstante, dado que para la generación del modelo geométrico en DesignBuilder solamente es necesario el perímetro de la edificación, se elaboró un plano arquitectónico más simple que contiene únicamente el perímetro mencionado. En este proceso se eliminaron los elementos complementarios que estaban en los planos originales, tal como se ilustra en la Figura 3.

Figura 3.

Plano de planta de piso del Edificio Qatar y construcción en DesignBuilder



Nota. Comparación entre el plano arquitectónico del piso y la geometría implementada en DesignBuilder

Una vez definidos ambos planos (apartamento y piso), el proceso de modelado se realizó en las siguientes etapas:

- Importación y modelado del apartamento: Se importó inicialmente el plano del apartamento, sobre el cual se construyó el primer bloque geométrico. A partir de este bloque se incorporaron los detalles internos correspondientes, tales como puertas, ventanas y divisiones.
- Modelado del piso completo: Posteriormente, se eliminó el plano importado del apartamento y se cargó el plano del piso. Mediante la herramienta mover de

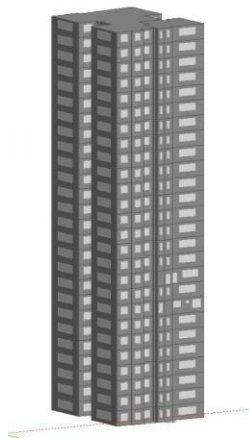
DesignBuilder, se ubicó el bloque del apartamento en la posición correcta, lo que permitió construir el bloque del resto del piso. Dado que la simulación energética se centra únicamente en el apartamento de estudio, a este bloque adicional no se le incluyeron detalles arquitectónicos.

- **Construcción del edificio:** Finalmente, se asumió que los 24 pisos del edificio poseen la misma configuración geométrica, utilizando el piso donde se ubica el apartamento como base de replicación para generar el modelo completo. Esta suposición se justifica en la medida en que, al enfocarse la simulación energética en el apartamento, las variaciones geométricas entre pisos no tienen un impacto significativo en los resultados obtenidos.

La Figura 4 presenta la construcción final del modelo geométrico, en la cual se integran los diferentes elementos desarrollados durante las etapas de modelado, incluyendo el apartamento, el piso completo y la configuración general del edificio. Esta representación permite visualizar de manera conjunta la geometría utilizada como base para la simulación energética.

Figura 4.

Modelo geométrico Edificio Qatar



Nota. Representación tridimensional del modelo geométrico desarrollado en DesignBuilder

4.3 Integración de datos en el modelo

En esta etapa se llevó a cabo la integración de los datos físicos y operativos del modelo energético, con el propósito de representar con precisión las condiciones constructivas de uso y de operación del apartamento en estudio. Este proceso permitió vincular la geometría

previamente modelada con la información necesaria para realizar simulaciones energéticas mediante el motor de cálculo EnergyPlus.

De esta manera, se procedió a editar los cerramientos y aberturas del apartamento, es decir, los materiales correspondientes a las paredes, puertas y ventanas.

Este ajuste tuvo como finalidad incrementar la precisión de la simulación energética, garantizando que las propiedades térmicas y constructivas reflejaran fielmente las condiciones reales del proyecto.

La información utilizada fue proporcionada por la empresa constructora Oteco, permitiendo definir los materiales y espesores de los diferentes tipos de muros del apartamento, los cuales se describen a continuación:

- Muro de fachada: conformado por ladrillo exterior, concreto liviano y bloque de concreto medio, con un espesor total de 0,22 m.

Figura 5.

Muro de fachada



Nota. Configuración constructiva del muro de fachada incorporado en el modelo energético

- Particiones internas generales: conformadas por yeso, mortero y ladrillo interior, con un espesor total de 0,12 m.

Figura 6.

Particiones internas generales



Nota. Definición de las particiones internas utilizadas en la simulación energética.

- División interna distinta (Apartamento de estudio): conformada por ladrillo, mortero y yeso, con un espesor total de 0,12 m.

Figura 7.

División interna distinta



Nota. Configuración constructiva de la división interna específica del apartamento de estudio.

Asimismo, con base en la información anteriormente mencionada, se establecieron las propiedades constructivas correspondientes a las puertas y ventanas del apartamento, con el fin de representar adecuadamente su comportamiento térmico dentro del modelo energético.

- Puertas: construidas en madera contrachapada (*Plywood Lightweight*) con cámara de aire intermedia de aproximadamente 25 mm, alcanzando un espesor total de 0,04 m. Esta configuración permite un peso reducido y un aislamiento térmico moderado, adecuado para interiores residenciales.

Figura 8.*Puertas internas del apartamento del Edificio Qatar*

Nota. Propiedades constructivas asignadas a las puertas internas del modelo energético

- Ventanas: compuestas por un único panel de vidrio tipo *Generic Clear* 4 mm, correspondiente a un acristalamiento simple transparente. Este tipo de vidrio permite una adecuada transmisión lumínica natural hacia los espacios interiores, con un comportamiento térmico estándar para edificaciones residenciales.

Finalmente, una vez incorporados todos los materiales correspondientes a los cerramientos y ventanas, se obtiene la vista renderizada final del edificio. Esta representación permite apreciar de manera visual la configuración del modelo y los materiales asignados, sirviendo como verificación del correcto modelado previo a la simulación energética. La vista renderizada resultante se presenta en la Figura 9.

Figura 9.

Vista renderizada del Edificio Qatar



Nota. Vista final del modelo tras la asignación de materiales y cerramientos.

Incorporación de cargas internas y perfiles de uso: se integraron cargas internas derivadas de la ocupación, iluminación y equipos eléctricos que se muestran en el Anexo A.0.1. Cada zona térmica fue configurada con sus respectivos perfiles horarios de operación, tomando como base la información real de funcionamiento del apartamento.

- Ocupación: Se estableció la densidad de personas ($\text{m}^2/\text{persona}$)
- Iluminación: En la sección de iluminación no se agregó información; esta se incluyó como carga en la sección de equipos.
- Equipos: Se integraron las cargas eléctricas asociadas a los equipos y sistemas, especificando sus horarios de funcionamiento para cada zona, mediante la configuración realizada en la pestaña “equipos” del software DesignBuilder.
- Ventilación e infiltración: Se configuraron las tasas de ventilación de acuerdo con los requerimientos de confort y calidad del aire interior.

Estos datos permitieron estimar la generación interna de calor y su variación en el tiempo, influyendo directamente en las cargas térmicas del edificio.

Condiciones ambientales y climáticas: Para la simulación se empleó un archivo climático EPW correspondiente a ciudad de Bucaramanga, Santander, el cual contiene información horaria de temperatura, humedad relativa, radiación solar, dirección y velocidad del viento.

Este archivo se integró al modelo para garantizar que las condiciones ambientales locales se representaran con precisión durante la simulación. Además, se definieron las condiciones de confort interior, estableciendo temperaturas de calefacción y refrigeración, y considerando estrategias de ventilación natural cuando las condiciones exteriores del edificio lo permitían.

Configuración del sistema HVAC: En esta pestaña se integraron los sistemas de refrigeración correspondientes a un equipo de aire acondicionado en el cuarto principal del apartamento, se ajustó la potencia nominal del equipo y la configuración de los horarios de uso. Lo que es ventilación natural, ventilación mecánica, control de humedad y calefacción no se tuvo en cuenta porque en este caso no aplica.

De manera complementaria, en DesignBuilder se define para cada equipo un perfil de operación que incluye tanto los horarios de uso como el nivel de utilización de su potencia nominal. Estos perfiles se especifican mediante porcentajes que representan la fracción de potencia demandada en cada intervalo de tiempo, tal como se muestra en el Anexo A.0.1. De esta forma, la simulación no solo indica cuándo un equipo está en funcionamiento, sino también cómo varía su consumo a lo largo del día. Por ejemplo, una nevera configurada como On (45 %) permanece activa durante todo el día, pero con un consumo promedio equivalente al 45 % de su potencia nominal, reflejando su funcionamiento cíclico y permitiendo una representación más realista del comportamiento energético del apartamento.

4.4 Ajuste del modelo

Teniendo en cuenta los equipos estimados y su correspondiente programación de uso, se realizó una primera simulación energética de la vivienda. Posteriormente, los resultados obtenidos fueron comparados con los consumos reales reportados por la Electrificadora de Santander (ESSA), con el fin de evaluar la precisión del modelo y su nivel de calibración. Los valores resultantes de esta comparación se presentan en la siguiente Tabla 2.

Tabla 2.

Simulación 1 en DesignBuilder

Mes	Simulado				Medido [kWh]	Error
	Equipos [kWh]	Iluminación [kWh]	Refrigeración [kWh]	Total [kWh]		
Enero	64,74	18,72	20,78	104,24	85	22,65%
Febrero	75,47	16,79	21,22	113,48	93	22,03%
Marzo	75,77	18,62	26,39	120,79	97	24,53%
Abril	73,27	17,95	28,60	119,82	91	31,68%
Mayo	84,34	18,62	33,14	136,10	101	34,76%
Junio	73,96	18,01	32,67	124,64	94	32,60%
Julio	76,80	18,56	35,03	130,39	91	43,29%
Agosto	83,29	18,62	29,81	131,72	107	23,11%
Septiembre	81,00	17,85	21,39	120,24	99	21,46%
Octubre	76,61	18,72	22,08	117,42	94	24,92%
Noviembre	80,81	18,01	21,09	119,91	98	22,37%
Diciembre	84,31	18,56	20,11	122,98	107	14,94%

Nota. Resultados iniciales del modelo antes del proceso de calibración.

De acuerdo con la *ASHRAE Guideline 14: Measurement of Energy, Demand, and Water Savings*, el modelo energético debe presentar un error inferior al 5 % entre los valores simulados y los datos medidos, con el fin de garantizar una calibración adecuada y una representación confiable del comportamiento energético real.

En la primera simulación, los porcentajes de error superaron este límite, lo que indicó la necesidad de ajustar los parámetros del modelo. Por ello, se realizaron modificaciones mensuales en los perfiles de uso y en la operación de los equipos, buscando mejorar la correspondencia entre los consumos simulados y los reales.

Adicionalmente, se eliminaron ciertos equipos inicialmente considerados, que posteriormente se identificó que no formaban parte del inventario real de la vivienda, según la información verificada con el propietario. Estos ajustes permitieron reducir progresivamente el error porcentual hasta alcanzar el rango aceptable establecido por la norma como se muestra en la Tabla 3.

Tabla 3.

Simulación 4 en DesignBuilder

Mes	Simulado			Total [kWh]	Medido	
	Equipos [kWh]	Iluminación [kWh]	Refrigeración [kWh]		[kWh]	Error
Enero	49,67	18,72	20,78	89,17	85	4,92%
Febrero	52,71	16,79	21,33	90,84	93	2,32%
Marzo	54,64	18,62	26,40	99,66	97	2,75%
Abril	52,04	17,95	22,18	92,17	91	1,29%
Mayo	53,88	18,62	25,51	98,02	101	2,95%
Junio	52,13	18,01	24,96	95,11	94	1,18%
Julio	53,79	18,56	17,67	90,02	91	1,07%
Agosto	56,66	18,62	29,90	105,19	107	1,69%
Septiembre	56,79	17,85	21,46	96,10	99	2,93%
Octubre	54,29	18,72	22,10	95,12	94	1,19%
Noviembre	56,55	18,01	21,14	95,70	98	2,34%
Diciembre	63,82	18,56	20,26	102,64	107	4,07%

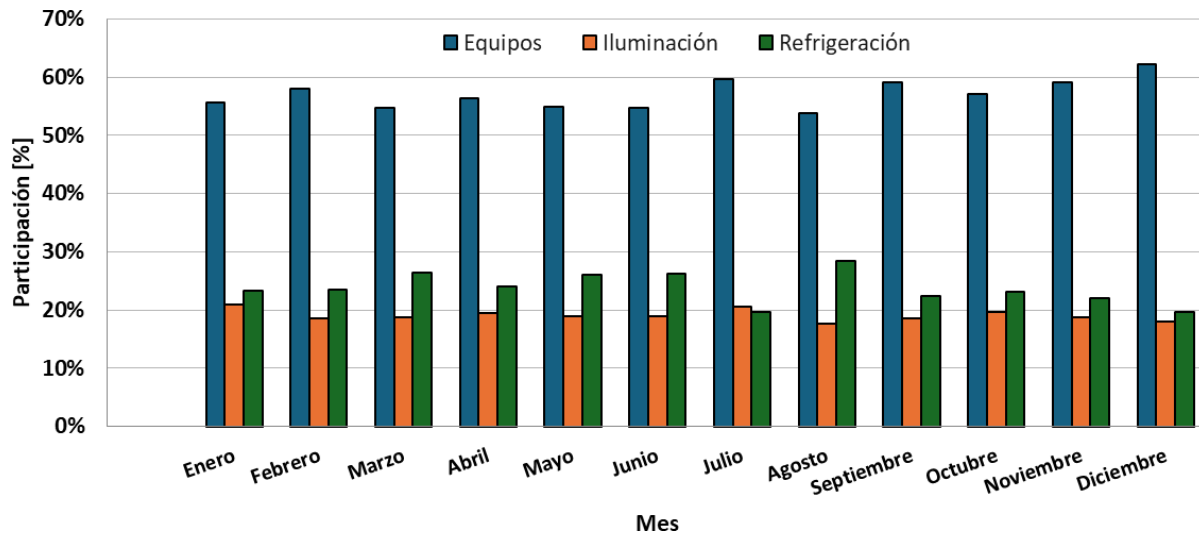
Nota. Resultados del modelo tras el ajuste de parámetros operativos.

Una vez calibrado el modelo y alcanzados los márgenes de error dentro de los límites establecidos por la *ASHRAE Guideline 14*, se observa que los equipos representan más de la mitad del consumo energético total de la vivienda durante todo el año, llegando a superar el 60 % en el mes de diciembre, tal como se aprecia en la Figura 10.

Este comportamiento evidencia que los equipos eléctricos constituyen la principal fuente de demanda energética, mientras que la iluminación y la refrigeración mantienen una participación relativamente constante y significativamente menor.

Figura 10.

Participación por uso final del apartamento



Nota. Distribución porcentual del consumo energético por uso final obtenida mediante simulación.

5. Modelo basado en datos

Esta sección desarrolla el modelo de predicción del consumo energético a partir de los datos medidos en la vivienda dispuestos en el *dataset*. Se describe el conjunto de datos, las variables disponibles y el proceso de preparación de la información, así como la implementación y evaluación de los distintos modelos de aprendizaje automático (Regresión Lineal, *Random Forest*, MLP y XGBoost). El objetivo es analizar el desempeño de cada enfoque para estimar el consumo real del apartamento y establecer una base de comparación entre ellos.

5.1 Descripción del *dataset*

El conjunto de datos utilizado en este trabajo proviene de una campaña experimental de monitorización realizada en cinco viviendas tipo apartamento de ingresos medios ubicadas en el Área Metropolitana de Bucaramanga. La campaña incluyó la medición simultánea de variables eléctricas, ambientales y operativas, asociadas tanto al uso de los espacios como al comportamiento de los ocupantes.

El consumo energético fue registrado a nivel de circuito eléctrico mediante sub-medición, permitiendo capturar con alto detalle la demanda de los principales usos finales. Las variables ambientales interiores (temperatura y humedad) y exteriores (temperatura, humedad y radiación solar) fueron medidas mediante sensores dedicados y una estación meteorológica, mientras que

la información operativa relacionada con ocupación y uso de espacios se obtuvo a partir de sensores de presencia y de apertura/cierre de puertas y ventanas.

Los datos fueron recolectados con una resolución temporal de 10 minutos para las variables ambientales y operativas, y de 5 minutos para el consumo energético, ajustándose progresivamente la instrumentación para priorizar los espacios con mayor incidencia en la demanda eléctrica.

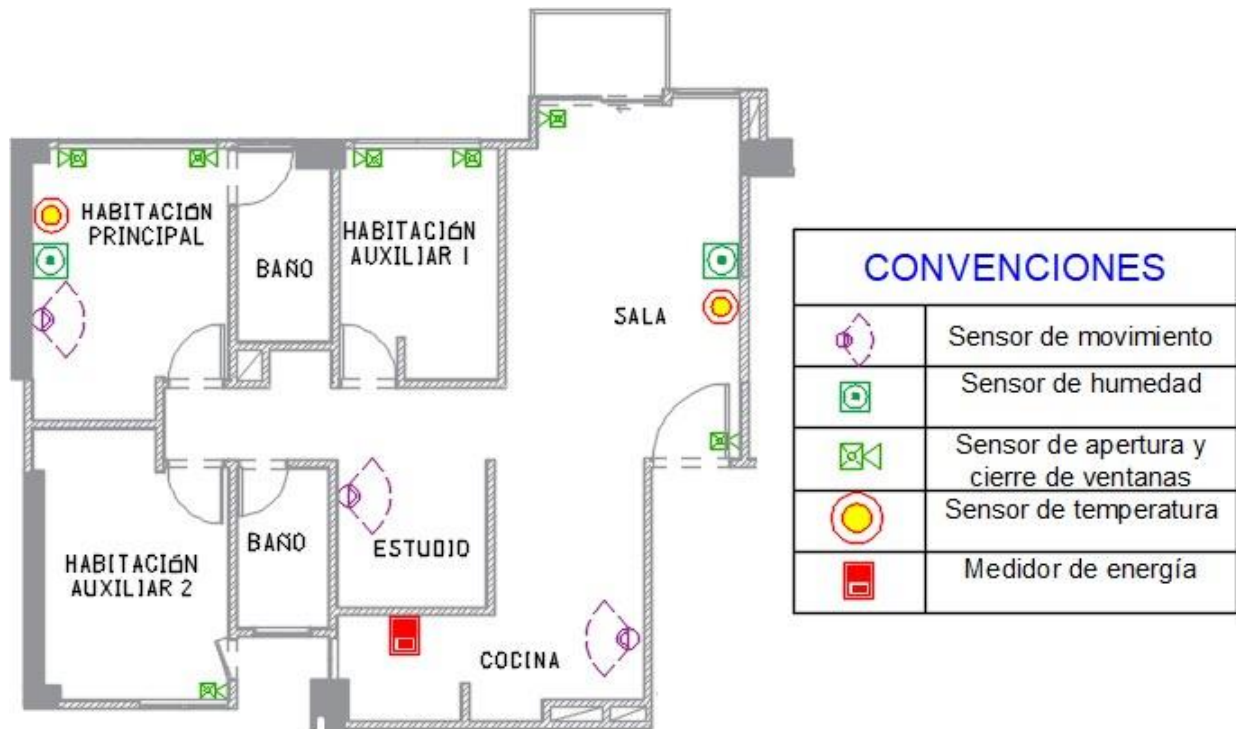
La submedición eléctrica se realizó mediante un medidor ACUREV 2110 (Accuenergy), utilizado para el registro del consumo eléctrico, la corriente y el factor de potencia del apartamento. La monitorización ambiental interior se efectuó mediante sensores ALTA MNS2-9-W2-TS-ST (IMONNIT), mientras que la detección de apertura y cierre de ventanas y de la puerta principal se realizó con sensores ALTA Wireless Open-Closed Sensors (IMONNIT). Adicionalmente, la presencia de ocupantes fue registrada a través de sensores ALTA Motion Detection Sensors (IMONNIT).

Las condiciones ambientales exteriores se midieron con una estación meteorológica Davis Vantage Pro 2, que proporcionó información de temperatura, humedad relativa y radiación solar.

El esquema de medición implementado en este estudio se encuentra alineado con la investigación doctoral en desarrollo titulada “Priorización de características de entrada en modelos de predicción de energía para viviendas bajo condiciones de clima cálido seco”, la cual sirve como marco conceptual de referencia. En este contexto, la Figura 11 presenta la dinámica general de medición adoptada en el presente trabajo, ilustrando la ubicación de los sensores, las variables registradas y su integración en la construcción del dataset utilizado para la evaluación de los modelos de predicción. El Anexo A.0.2 resume y describe detalladamente las variables incluidas en el dataset.

Figura 11.

Ubicación de sensores y equipo de medida de energía



Nota. Esquema de instrumentación empleado durante la campaña experimental de monitoreo.

5.2 Revisión y descripción de los modelos de predicción

De acuerdo con la naturaleza del dataset descrito en la Sección 5.1, el problema de estimación del consumo energético en este estudio no resulta adecuadamente representado mediante un enfoque exclusivamente lineal. El consumo eléctrico en una vivienda depende de múltiples factores que interactúan entre sí, como las condiciones ambientales, la ocupación de los espacios, el uso de equipos, los horarios y las dinámicas propias de los habitantes. Estas interacciones generan comportamientos complejos y variaciones no lineales a lo largo del tiempo.

Este tipo de comportamiento ha sido ampliamente reconocido en la literatura, donde se señala que la predicción del consumo energético en edificios y viviendas requiere modelos capaces de capturar tanto relaciones lineales como no lineales, así como dependencias temporales inherentes a la serie de consumo (Lahouar & Slama, 2015).

Antes de definir los modelos finales, se evaluaron diferentes algoritmos de Machine Learning reportados en estudios previos sobre predicción de consumo energético residencial, entre ellos: Regresión Lineal (RL), Árboles de Decisión (DT), *Random Forest* (RF), Máquinas de Vectores de Soporte (SVM) y Redes Neuronales (RN), de acuerdo con la clasificación presentada

en el documento de selección de modelos.

Dada la naturaleza temporal de los datos, la validación de los modelos se llevó a cabo mediante una partición estrictamente temporal, reservando el último día completo de medición como conjunto de prueba y utilizando el resto del histórico para el entrenamiento. Este esquema, conocido en la literatura como last day hold-out, es ampliamente utilizado en problemas de predicción de demanda energética, ya que evita la fuga de información asociada a particiones aleatorias y permite evaluar el desempeño del modelo respetando la secuencia temporal de los datos (Hyndman & Athanasopoulos, 2021; Hong, Pinson, & Fan, 2014).

Por otra parte, el tamaño del *dataset* utilizado en este trabajo aproximadamente dos meses de medición con una resolución temporal de 10 minutos implica un conjunto de datos de tamaño medio, suficiente para entrenar modelos estadísticos y de complejidad moderada, pero limitado para redes neuronales profundas que requieren largos históricos temporales. Adicionalmente, al tratarse de una única vivienda, existe un riesgo importante de sobreajuste en modelos excesivamente complejos. Por estas razones, se seleccionaron finalmente tres modelos para el desarrollo de la predicción: Regresión Lineal (RL), *Random Forest* (RF) y Redes Neuronales Multicapa (MLP).

- **Regresión Lineal (RL)**

La regresión lineal es uno de los modelos más utilizados en estudios de predicción debido a su simplicidad y facilidad de interpretación. Su objetivo es establecer una relación matemática directa entre el consumo energético y un conjunto de variables explicativas, tales como la temperatura, la hora del día, el tipo de día y el uso de determinados equipos. En este modelo, la variable objetivo se expresa como una combinación lineal de las variables de entrada, de acuerdo con la ecuación (1):

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \varepsilon \quad (1)$$

Donde y representa el consumo energético, x_1, x_2, \dots, x_n son las variables explicativas, β_0 es el término independiente, $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes que cuantifican la influencia de cada variable, y ε corresponde al término de error. Durante el proceso de entrenamiento, el modelo estima estos coeficientes minimizando la suma de los errores cuadrados entre los valores reales y los valores predichos, lo que permite obtener la mejor aproximación lineal posible del comportamiento observado.

Diversos estudios han utilizado la regresión lineal como modelo de referencia en problemas de predicción energética. En general, se ha reportado que este enfoque puede ofrecer resultados aceptables en escenarios donde el comportamiento del consumo es relativamente estable y las relaciones entre variables presentan una estructura predominantemente lineal; sin embargo, su desempeño tiende a disminuir cuando existen interacciones complejas o relaciones no lineales marcadas (Lahouar & Slama, 2015).

De manera similar, Deb et al. (2017) emplearon la regresión lineal como modelo base para comparar el desempeño de técnicas más avanzadas en la predicción del consumo energético en edificaciones, destacando su utilidad como punto de referencia, aunque con limitaciones frente a modelos no lineales.

En este trabajo, la regresión lineal se emplea como un modelo de referencia, ya que permite cuantificar de manera clara la mejora obtenida al aplicar modelos no lineales. No obstante, su capacidad para representar la estacionalidad y la complejidad del consumo es limitada, por lo que su desempeño esperado se encuentra entre bajo y medio.

- ***Random Forest (RF)***

El modelo Random Forest es un algoritmo de aprendizaje no lineal basado en un conjunto de árboles de decisión, propuesto por Breiman (2001). Como se ilustra en la Figura 11, este modelo se fundamenta en la construcción de múltiples árboles entrenados de forma independiente a partir del mismo conjunto de datos, pero utilizando distintos subconjuntos de muestras y variables mediante técnicas de bootstrap y selección aleatoria de características. Cada árbol genera una predicción individual y, posteriormente, dichas predicciones se combinan a través de un proceso de promediado, obteniendo así el resultado final. Este enfoque permite reducir la varianza del modelo y mejorar la estabilidad y precisión frente a un único árbol de decisión.

En el ámbito energético, Random Forest ha sido ampliamente aplicado y validado en problemas de predicción de demanda. Lahouar y Slama (2015) reportaron resultados competitivos en el pronóstico de carga a corto plazo, mientras que Ahmad et al. (2018) lo emplearon en la predicción del consumo energético en edificaciones, destacando su capacidad para capturar relaciones no lineales entre variables climáticas, ocupación y horarios de uso.

Random Forest resulta especialmente adecuado para el presente proyecto, dado que el patrón de consumo del apartamento depende de interacciones complejas entre múltiples variables

medidas, tales como sensores de movimiento, apertura de ventanas, uso de equipos eléctricos y condiciones ambientales. Además, una de sus principales ventajas es la posibilidad de estimar la importancia relativa de las variables, lo que facilita la identificación de los factores con mayor influencia sobre el consumo energético.

- **Redes Neuronales Multicapa (MLP)**

En este trabajo se decidió utilizar Redes Neuronales Multicapa (MLP) como uno de los modelos de predicción. Este tipo de red pertenece a la familia de las redes neuronales artificiales y se caracteriza por estar compuesta por múltiples capas de neuronas interconectadas, las cuales transforman progresivamente la información desde la capa de entrada hasta la capa de salida, donde se obtiene la predicción final. La Figura 12 presenta un esquema general de la arquitectura de una MLP, ilustrando la capa de entrada, las capas ocultas y la capa de salida, así como las conexiones ponderadas entre ellas.

A diferencia de los modelos lineales, las MLP tienen la capacidad de aprender relaciones no lineales complejas entre las variables de entrada. Esta característica resulta especialmente útil en el presente estudio, dado que el consumo energético no depende únicamente del tiempo, sino también de factores como la ocupación del apartamento, el uso de equipos eléctricos, la ventilación y las condiciones ambientales, los cuales interactúan dinámicamente a lo largo del día.

Durante el proceso de entrenamiento, la red ajusta sus parámetros internos mediante algoritmos de optimización, con el objetivo de minimizar el error entre los valores predichos y los valores reales de consumo. Cuando la arquitectura y los hiperparámetros son adecuadamente seleccionados, la red puede identificar patrones subyacentes que no son evidentes mediante modelos lineales tradicionales. Diversos estudios han reportado resultados favorables en la predicción de carga eléctrica utilizando redes neuronales, siempre que se controle el sobreajuste y se mantenga un equilibrio entre complejidad del modelo y cantidad de datos disponibles (Hippert et al., 2001).

En el contexto de este proyecto, las MLP representan una alternativa interesante, ya que permiten modelar múltiples variables simultáneamente y capturar interacciones no lineales presentes en el dataset. No obstante, dado que el conjunto de datos disponible abarca aproximadamente dos meses de medición, resulta necesario evitar arquitecturas excesivamente profundas o con un número elevado de parámetros, ya que podrían ajustarse de manera excesiva

al comportamiento específico del apartamento y reducir su capacidad de generalización.

- **XGBoost**

El modelo XGBoost (Extreme Gradient Boosting) es un algoritmo de aprendizaje no lineal basado en el enfoque de boosting mediante árboles de decisión, propuesto por Chen y Guestrin (2016). Su funcionamiento se fundamenta en la construcción secuencial de árboles, donde cada nuevo modelo se entrena con el objetivo de corregir los errores del conjunto anterior, optimizando el proceso de aprendizaje a partir del gradiente de la función de pérdida.

En el ámbito energético, XGBoost ha sido ampliamente aplicado debido a su capacidad para modelar relaciones no lineales complejas y manejar conjuntos de datos con múltiples variables. Ahmad et al. (2018) reportaron resultados competitivos en la predicción del consumo energético en edificaciones, destacando su desempeño frente a otros modelos de aprendizaje automático. De manera similar, Barzola-Monteses et al. (2021) emplearon XGBoost en la estimación del consumo energético de edificaciones, evidenciando su capacidad para integrar variables asociadas al comportamiento de los usuarios y a las condiciones operativas del edificio.

En el contexto de este proyecto, XGBoost constituye una alternativa adecuada debido a su capacidad para capturar interacciones complejas entre las variables medidas —como ocupación, uso de equipos eléctricos y condiciones ambientales—, permitiendo representar de forma eficiente el comportamiento dinámico del consumo energético del apartamento a lo largo del día.

5.3 Elaboración del modelo

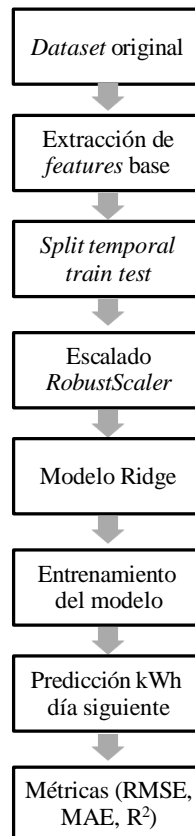
El desarrollo del modelo se realizó con base en la revisión previa presentada en la Sección 5.2, a partir de la cual se implementaron cuatro modelos de predicción: Regresión lineal, *Random Forest*, MLP y XGBoost. Todos los modelos se trabajaron bajo el mismo esquema general de procesamiento de datos y evaluación, con el objetivo de compararlos en igualdad de condiciones. A continuación, se presentan los diagramas de bloques que muestran de manera general cómo se llevó a cabo la implementación de cada uno de los modelos.

5.3.1 Implementación del modelo de Regresión Lineal

A continuación, se presenta la estructura básica de la implementación del modelo de Regresión Lineal.

Figura 12.

Implementación del modelo de Regresión Lineal.



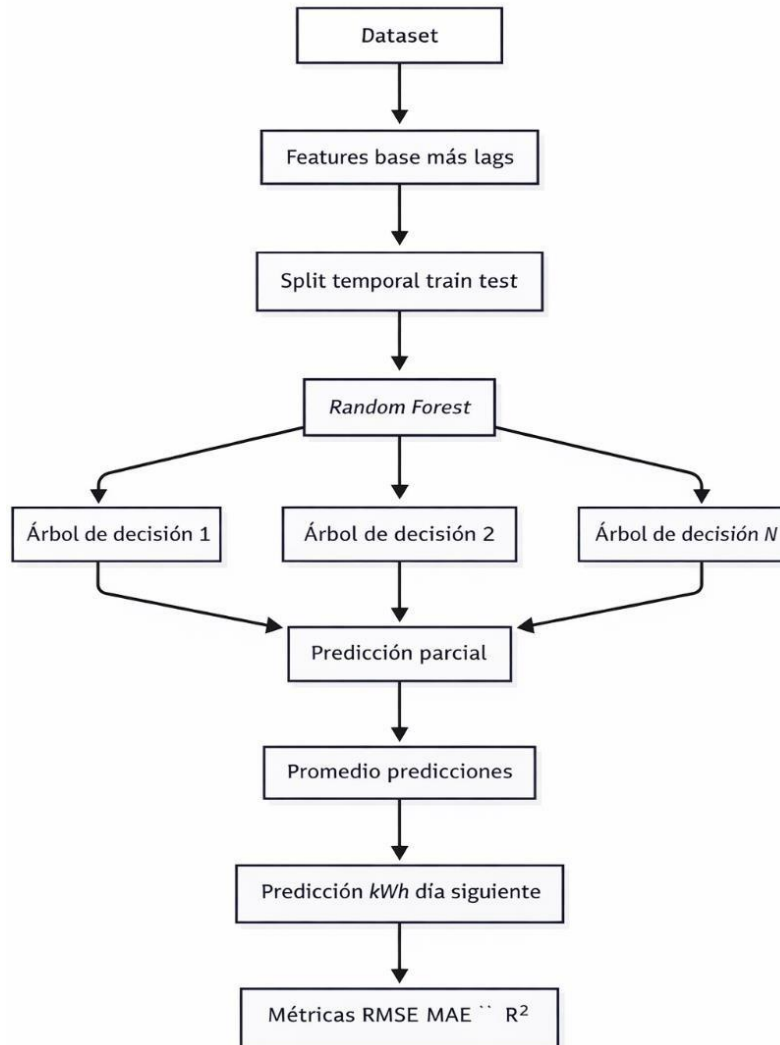
Nota. Diagrama de bloques del proceso de entrenamiento y predicción del modelo de Regresión Lineal.

5.3.2 Implementación del modelo de Random Forest

A continuación, se presenta la estructura básica de la implementación del modelo de *Random Forest*.

Figura 13.

Implementación del modelo de Random Forest



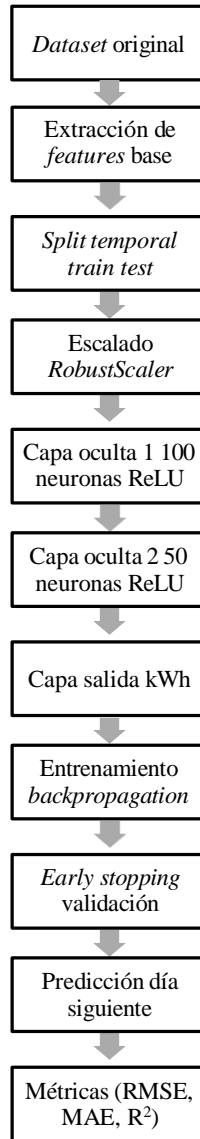
Nota. Estructura general del modelo Random Forest utilizada para la predicción del consumo energético.

5.3.3 Implementación del modelo de MLP

A continuación, se presenta la estructura básica de la implementación del modelo de Redes Neuronales Multicapa (MLP).

Figura 14.

Implementación del modelo de MLP



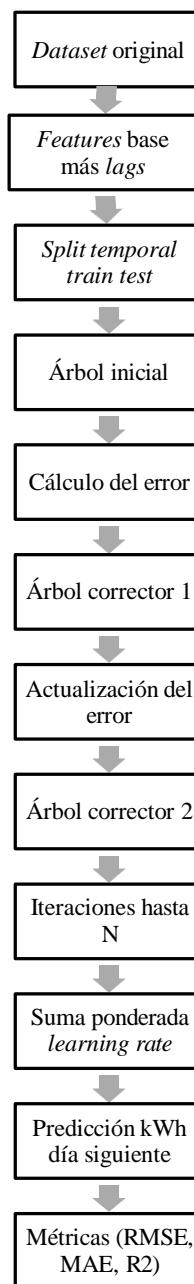
Nota. Arquitectura de la red neuronal multicapa implementada en el estudio.

5.3.4 Implementación del modelo de XGBoost

A continuación, se presenta la estructura básica de la implementación del modelo de XGBoost

Figura 15.

Implementación del modelo de XGBoost



Nota. Esquema conceptual del proceso iterativo de boosting aplicado en XGBoost.

5.4 Evaluación de rendimiento de los modelos

Para la evaluación integral del desempeño de los modelos se consideraron, en primer lugar, criterios asociados a su demanda computacional, tales como el tiempo de entrenamiento y el tiempo de predicción, con el objetivo de analizar su viabilidad operativa y su aplicabilidad en contextos de predicción rápida. Los resultados correspondientes a la medición del tiempo de entrenamiento y del tiempo de predicción se presentan en la Tabla 4.

Tabla 4.

Resultados de demanda computacional de los modelos basados en datos

Modelo	Tiempo entrenamiento	
	[s]	Tiempo predicción [s]
Regresión		
Lineal	0,0104	0,0003
MLP	10,0318	0,0005
<i>Random Forest</i>	34,9124	0,2635
XGBoost	1,2926	0,0097

Nota. Tiempos de entrenamiento y predicción de los modelos evaluados.

Como se puede observar, el modelo *Random Forest* es el que presenta la mayor demanda computacional, tanto en la etapa de entrenamiento como en la fase de predicción. Este comportamiento se explica por el funcionamiento del algoritmo, el cual se basa en la construcción de un elevado número de árboles de decisión independientes, así como en la profundidad alcanzada por estos. Durante el proceso de entrenamiento, cada árbol se ajusta utilizando subconjuntos de datos y variables, lo que incrementa el costo computacional total.

No obstante, dado que el *dataset* empleado en este estudio es de tamaño moderado, las diferencias observadas en la demanda computacional no representan una limitación significativa en el contexto analizado. Estas diferencias podrían volverse relevantes en escenarios con conjuntos de datos de mayor tamaño o en aplicaciones que requieran horizontes de predicción más extensos que un día, donde el costo computacional tendría un impacto más considerable.

Para la evaluación del desempeño de los modelos se seleccionaron cuatro métricas, con el fin de realizar una comparación objetiva, consistente y equitativa entre ellos y, a partir de ello, extraer conclusiones fundamentadas sobre su capacidad predictiva. Las métricas consideradas fueron el error cuadrático medio (RMSE), que penaliza con mayor peso los errores de gran magnitud; el error absoluto medio (MAE), que cuantifica el error promedio en términos absolutos; el coeficiente de determinación (R^2), que indica la proporción de la variabilidad del consumo

explicada por el modelo; y el error porcentual absoluto medio simétrico (SMAPE), el cual expresa el error en términos porcentuales normalizándolo mediante el promedio de las magnitudes absolutas del valor real y del valor estimado. Esta métrica fue seleccionada en lugar del MAPE tradicional debido a su menor sensibilidad a valores reales cercanos a cero, condición frecuente en la predicción del consumo energético residencial, lo que permite obtener una evaluación más estable y representativa del desempeño de los modelos. A continuación, se presenta la formulación matemática de cada una de estas métricas.

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

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \tag{4}$$

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \tag{5}$$

Esta selección se sustenta en la revisión de investigaciones previas, en las cuales dichas métricas se emplean de manera recurrente y han demostrado ser efectivas para la evaluación y comparación del desempeño de modelos de predicción. Los resultados obtenidos para la estimación del consumo en el día de prueba se presentan en la Tabla 5. Asimismo, la Figura 16 muestra la curva de consumo estimada por cada uno de los cuatro modelos a lo largo del día completo.

Tabla 5.

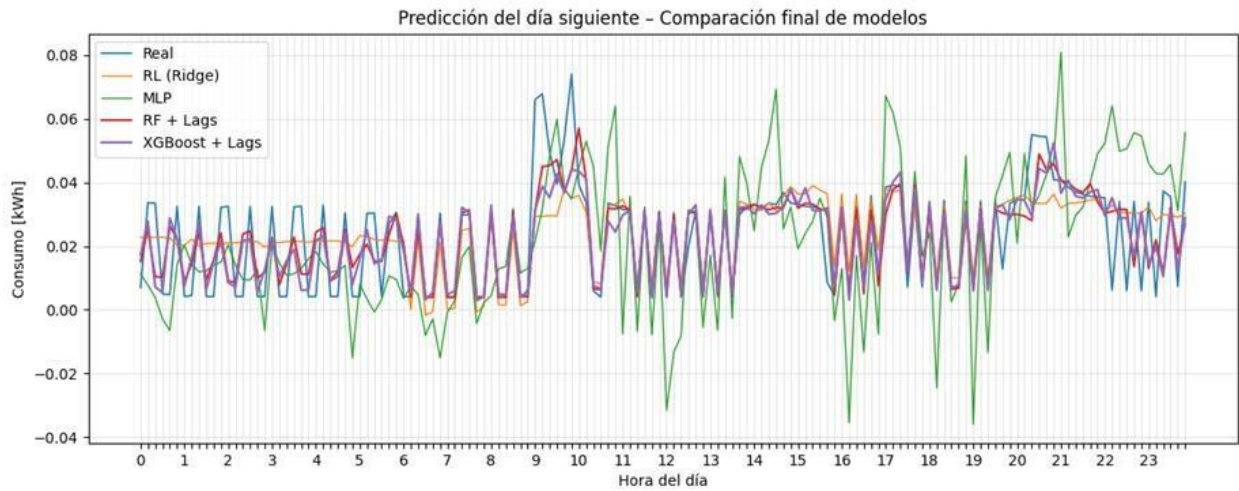
Resultados de los modelos basados en datos

Modelo	RMSE [kWh]	MAE [kWh]	R ²	SMAPE
Regresión Lineal	0,0118	0,0085	0,4888	112%
MLP	0,0198	0,0158	-0,4365	128%
<i>Random Forest</i>	0,0101	0,0065	0,6219	94%
XGBoost	0,0104	0,0067	0,6054	91%

Nota. Métricas de desempeño obtenidas en el día de prueba.

Figura 16.

Resultados de los modelos basados en datos



Nota. Comparación de las curvas de consumo predichas por los modelos evaluados.

Observando los resultados presentados en la Tabla 5 y el comportamiento de las curvas en la Figura 16, el modelo MLP muestra la capacidad de capturar relaciones no lineales entre las variables de entrada y el consumo energético, lo cual en principio constituye una ventaja frente a modelos puramente lineales. Además, el modelo logra reproducir parcialmente la tendencia general del consumo durante el día analizado.

No obstante, al analizar detalladamente su desempeño, se evidencia que el modelo MLP presenta las peores métricas entre todos los modelos evaluados, con un RMSE de 0,0198 kWh, un MAE de 0,0158 kWh, un SMAPE de 128 % y un coeficiente de determinación $R^2 = -0,4365$. Adicionalmente, en la representación gráfica se observan picos con valores negativos y una alta inestabilidad los valores de la predicción, lo que indica una pobre capacidad de generalización y una limitada confiabilidad para este caso de estudio. Por estas razones, el modelo MLP se descarta como una alternativa adecuada para la predicción del consumo energético en el escenario analizado.

Este comportamiento se explica por la naturaleza del modelo, ya que las redes neuronales suelen requerir conjuntos de datos más amplios para lograr un entrenamiento adecuado. Al contar con un número elevado de parámetros que deben ajustarse durante el proceso de aprendizaje, un *dataset* limitado restringe la capacidad del modelo para aprender patrones representativos del fenómeno estudiado. Esta limitación es especialmente relevante en la predicción del consumo energético, un proceso altamente variable en el que influyen múltiples

factores que cambian a lo largo del tiempo, como la ocupación, el uso de equipos y las condiciones ambientales.

En segundo lugar, se encuentra la regresión lineal, la cual, si bien presenta un desempeño superior al de la MLP, sigue siendo inferior al de los modelos basados en árboles de decisión. Este modelo alcanza un RMSE de 0,0118 kWh, un MAE de 0,0085 kWh, un SMAPE de 112 % y un $R^2 = 0,4888$. Su desempeño limitado se debe principalmente a que la regresión lineal asume una relación lineal entre las variables de entrada y la variable objetivo, suposición que rara vez se cumple en sistemas reales de consumo energético residencial.

En un apartamento, el consumo energético está influenciado por múltiples variables que interactúan de manera no lineal, tales como la ocupación, el uso simultáneo de equipos eléctricos, los cambios en la ventilación, la temperatura y los hábitos de los usuarios. Estas interacciones suelen presentar comportamientos abruptos, saturaciones y umbrales que no pueden ser representados adecuadamente mediante un modelo lineal.

Finalmente, la comparación entre *Random Forest* y XGBoost muestra un desempeño notablemente superior frente a los demás modelos, lo cual se explica principalmente por su capacidad para modelar relaciones no lineales y capturar interacciones complejas entre las variables de entrada. Al estar basados en conjuntos de árboles de decisión, ambos modelos no imponen una forma funcional fija y son capaces de adaptarse mejor a la variabilidad inherente del consumo energético residencial. En este contexto, *Random Forest* presenta los mejores resultados globales, con el menor RMSE (0,0101 kWh) y MAE (0,0065 kWh), así como el mayor coeficiente de determinación ($R^2 = 0,6219$), lo que indica una mayor capacidad para capturar la tendencia general y la variabilidad del consumo energético. Por su parte, XGBoost obtiene métricas muy cercanas, con un RMSE de 0,0104 kWh, un MAE de 0,0067 kWh, un $R^2 = 0,6054$ y el menor SMAPE (91 %), lo que sugiere una mayor estabilidad relativa en términos porcentuales.

A pesar del buen desempeño de ambos modelos, el mayor valor de R^2 y los menores errores absolutos obtenidos por *Random Forest* permiten concluir que este modelo es el que presenta el mejor desempeño global para la predicción del consumo energético del apartamento bajo las condiciones analizadas, siendo capaz de representar de forma más precisa la variabilidad y el comportamiento dinámico del consumo.

6. Comparación de modelos de predicción

En esta sección se presenta la comparación entre el modelo de predicción basado en datos y el modelo de ingeniería desarrollado en DesignBuilder, utilizando como referencia el consumo real del apartamento. A partir de esta comparación se analizan las diferencias en el comportamiento de las curvas de consumo, así como las métricas de error obtenidas por cada enfoque. El propósito de este análisis es determinar en qué medida cada modelo es capaz de reproducir el comportamiento real del consumo energético de la vivienda y qué tan confiables resultan sus predicciones en condiciones reales de operación. A continuación, se presenta la comparación de los modelos y el error de estos.

6.1 Comparación de los modelos

Para la comparación final con DesignBuilder, se seleccionó el modelo de *machine learning* con mejor desempeño global, correspondiente a *Random Forest*. Con el fin de realizar una comparación coherente y equitativa, se utilizaron los datos de consumo energético correspondientes a la predicción generada por DesignBuilder para el mismo día empleado en la evaluación de los modelos basados en datos, es decir, el 30 de noviembre, manteniendo una resolución temporal de 10 minutos.

Antes de abordar el análisis cuantitativo del desempeño predictivo, resulta pertinente comparar la demanda computacional asociada a cada enfoque, dado que DesignBuilder y los modelos de *machine learning* responden a filosofías de modelado fundamentalmente distintas. Esta comparación permite contextualizar los resultados posteriores no solo en términos de precisión, sino también de viabilidad operativa y esfuerzo computacional requerido.

Desde el punto de vista computacional, DesignBuilder presenta una elevada demanda durante cada ejecución de la simulación, ya que resuelve de manera explícita los balances energéticos y térmicos del edificio mediante modelos físicos detallados. Esta demanda aumenta con la complejidad geométrica del modelo, el número de zonas térmicas y la resolución temporal considerada, lo que implica que cada simulación constituye un proceso computacionalmente costoso, incluso para horizontes de predicción relativamente cortos.

En contraste, el modelo *Random Forest* concentra su mayor costo computacional en la etapa de entrenamiento, asociada a la construcción de múltiples árboles de decisión. No obstante, una vez entrenado, el tiempo requerido para realizar predicciones es considerablemente menor, lo que lo convierte en una alternativa eficiente para aplicaciones de

predicción operativa del consumo energético, especialmente cuando se requieren evaluaciones repetitivas o en tiempo casi real.

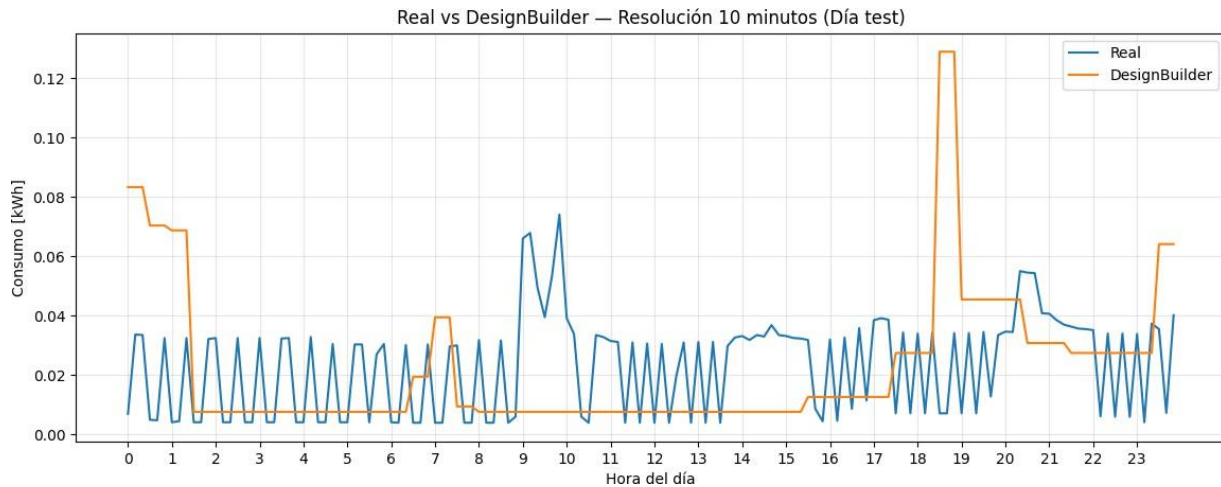
Adicionalmente, las entradas requeridas por cada enfoque también influyen en su carga computacional y en la complejidad de implementación. DesignBuilder requiere información detallada de tipo geométrico, climático y, especialmente, de la programación de uso de los equipos y ocupación, lo cual implica un mayor esfuerzo en la construcción y calibración del modelo. Por su parte, el enfoque basado en datos depende de la disponibilidad de un *dataset* representativo; en este estudio se utilizaron aproximadamente dos meses de mediciones, aunque en otros contextos este requerimiento puede ser significativamente mayor, incrementando a su vez la demanda computacional del proceso de entrenamiento.

A pesar de estas diferencias en términos de carga computacional, la comparación del desempeño predictivo puede efectuarse dentro de un marco común de evaluación. Aunque DesignBuilder no corresponde a un modelo predictivo entrenable, sino a una herramienta de simulación energética basada en principios físicos, la serie temporal de consumo estimado obtenida a partir de la simulación permite aplicar las mismas métricas de evaluación empleadas para los modelos de *machine learning*. En este sentido, el consumo simulado por DesignBuilder fue comparado directamente con el consumo real medido, calculándose métricas como RMSE, MAE, R^2 y SMAPE, las cuales cuantifican la discrepancia entre la simulación y la medición real bajo un mismo criterio de análisis.

La Figura 17 presenta la comparación entre las curvas de consumo real del día seleccionado y la estimación obtenida mediante DesignBuilder.

Figura 17.

Curvas de consumo Real, DesignBuilder



Nota. Comparación entre el consumo medido y el consumo estimado mediante simulación energética.

Como se puede observar, a pesar de que se aplicó el ajuste propuesto por la *ASHRAE Guideline 14: Measurement of Energy, Demand, and Water Savings*, logrando mantener el error de las mediciones mensuales por debajo del 5 %, el desempeño de la predicción resulta deficiente al compararse con la curva de consumo real. Esta discrepancia se evidencia tanto en el comportamiento de la curva como en las métricas obtenidas por el modelo, las cuales se presentan en la Tabla 6.

Tabla 6.

Métricas de comparación de DesignBuilder

	RMSE	MAE		
Modelo	[kWh]	[kWh]	R ²	SMAPE
DesignBuilder	0,0301	0,0217	-2,3380	62%

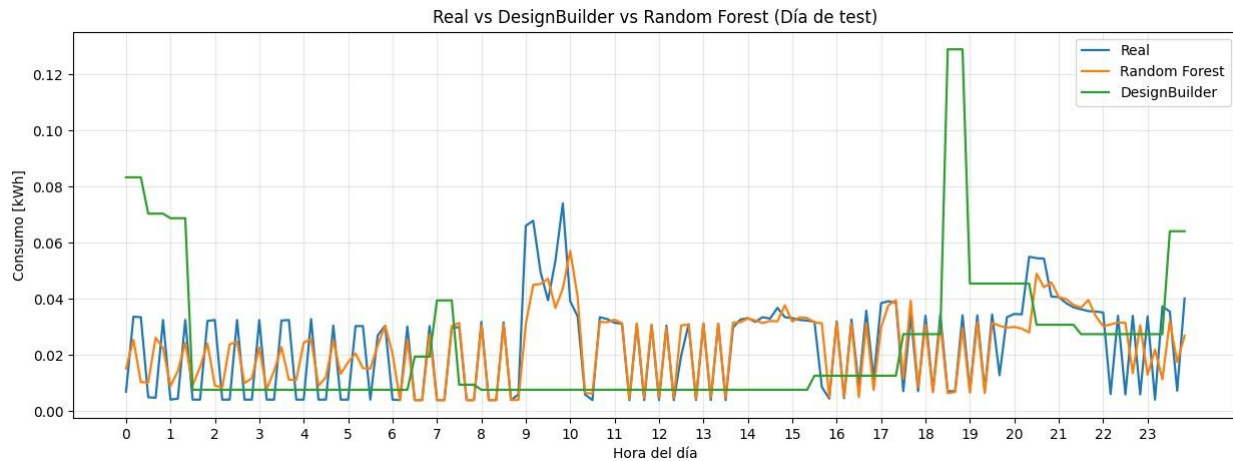
Nota. Evaluación del modelo físico frente al consumo medido.

No obstante, a pesar de que DesignBuilder presenta un valor relativamente bajo de SMAPE (62 %), se observa que el modelo no replica adecuadamente los picos presentes en la curva de consumo real. Esto se debe a que el SMAPE no penaliza de forma adecuada los errores asociados a valores extremos, a diferencia de métricas como el RMSE y el MAE, en las cuales DesignBuilder obtiene los peores resultados (RMSE = 0,0301 kWh y MAE = 0,0217 kWh). Adicionalmente, el coeficiente de determinación obtenido ($R^2 = -2,3380$) indica que el modelo presenta un desempeño inferior incluso al de una predicción basada en el valor medio. En

conjunto, estos resultados evidencian que DesignBuilder no logra capturar la variabilidad real del consumo energético, por lo que no es capaz de reproducir de manera adecuada ni la tendencia general ni los picos de consumo. En este contexto, el bajo valor de SMAPE resulta engañoso como indicador del desempeño real del modelo.

Figura 18.

Curvas de consumo Real, Random Forest y DesignBuilder



Nota. Comparación del desempeño del modelo Random Forest frente al modelo físico.

En la Figura 18 se observa que la curva de DesignBuilder presenta un comportamiento sobredimensionado durante la hora pico de consumo, correspondiente al intervalo de 6:00 a 7:00 p. m., y, de manera opuesta, se encuentra infra-dimensionada durante las horas valle. Este comportamiento evidencia que DesignBuilder es una herramienta adecuada para realizar estimaciones globales de consumo energético; sin embargo, al emplearse con fines de predicción a corto plazo, presenta limitaciones para reproducir con precisión las tendencias temporales y los picos de consumo que se manifiestan a lo largo de un día.

Por otro lado, el modelo de *Random Forest*, si bien no reproduce el comportamiento real de manera perfecta, logra capturar de forma más acertada tanto las tendencias generales como los picos de consumo, mostrando un desempeño superior en comparación con su contraparte basada en simulación física.

6.2 Error en los modelos

En la Figura 19 se muestra el error punto a punto del modelo basado en datos *Random Forest* y DesignBuilder para un día cualquiera, considerando datos de consumo cada 10 minutos y comparándolos con el consumo real del apartamento. Se observa que la variación del error es

considerable, con picos muy altos a lo largo del día, lo que indica que los modelos tienen dificultades para ajustarse de forma precisa al comportamiento real del consumo energético del apartamento.

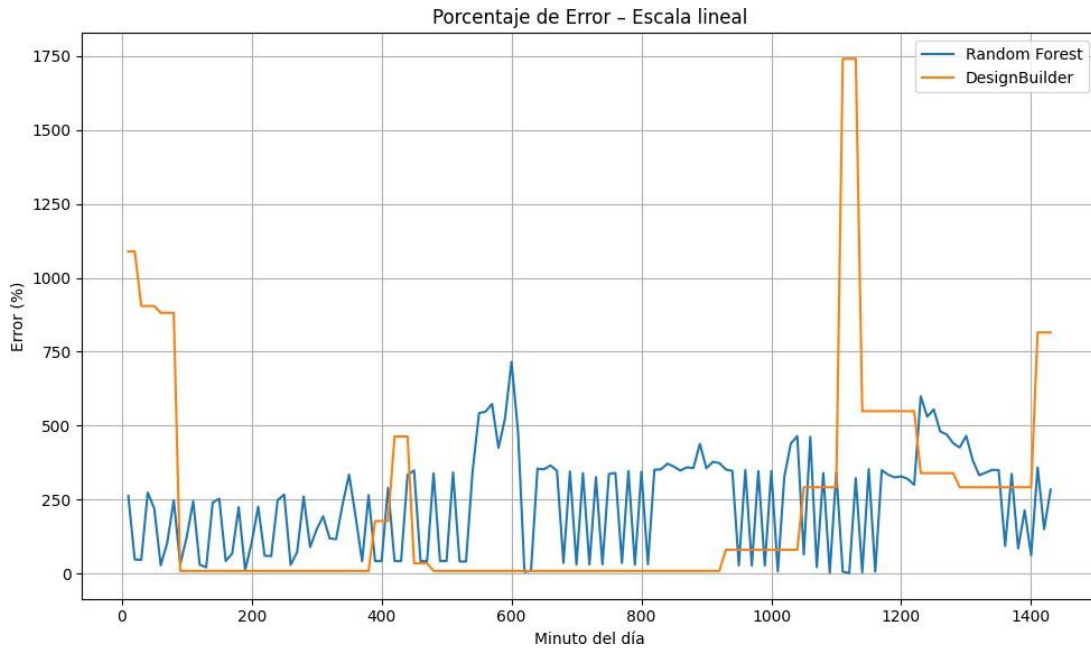
Esta alta variabilidad se explica principalmente porque el conjunto de datos disponible es reducido, lo que limita la capacidad de los modelos para aprender patrones representativos del consumo energético de la vivienda. Como es un entorno residencial, el consumo cambia bruscamente por el encendido o apagado de equipos, lo que provoca errores elevados cuando estos eventos no han sido previamente aprendidos por el modelo.

Al comparar los resultados, se evidencia que *Random Forest* presenta los menores errores, mostrando una mejor capacidad para adaptarse a estas variaciones, mientras que *DesignBuilder* registra los errores más altos, ya que se basa en supuestos de operación más generales y no captura con precisión los hábitos reales de los ocupantes. Sin embargo, también se identifican valles de error cercanos a cero, los cuales coinciden con intervalos de bajo consumo energético, donde el modelo de simulación tiende a estimar consumos prácticamente nulos. En estos casos, la diferencia absoluta entre el valor simulado y el valor medido es pequeña, lo que se traduce en errores porcentuales reducidos.

Este comportamiento confirma que la interpretación del error porcentual debe realizarse con cautela, ya que valores bajos de consumo amplifican artificialmente el error relativo, incluso cuando las desviaciones absolutas son pequeñas. En conjunto, la gráfica pone en evidencia la necesidad de contar con series de datos más extensas para reducir la dispersión del error y mejorar la confiabilidad de las predicciones.

Figura 19.

Porcentaje de error entre modelo de Random Forest, DesignBuilder



Nota. Error porcentual entre el consumo real y las predicciones de los modelos analizados.

7. Conclusiones

En este proyecto se desarrolló y evaluó un modelo de predicción del consumo energético de una vivienda de ingresos medios-altos, utilizando un enfoque basado en datos reales y comparándolo con un modelo de ingeniería desarrollado mediante simulación. Los resultados obtenidos permiten extraer varias conclusiones relevantes sobre el desempeño y las limitaciones más importantes de cada enfoque.

En primer lugar, se evidenció que el consumo energético residencial presenta un comportamiento altamente dinámico y no lineal, influenciado por múltiples factores como la ocupación, el uso simultáneo de electrodomésticos y demás equipos eléctricos, la ventilación natural y las condiciones ambientales. Por este motivo, esta complejidad dificulta la predicción precisa del consumo, especialmente cuando se trabaja con series de datos relativamente cortas, como en este caso, al disponer de aproximadamente dos meses de medición.

Los modelos basados en datos demostraron una mayor capacidad para adaptarse a la variabilidad real del consumo energético de la vivienda que el modelo de ingeniería. En particular, el modelo de *Random Forest* mostró el mejor desempeño global, logrando representar de manera

más consistente tanto la tendencia general como las variaciones del consumo de energía a lo largo del día. Esto se debe principalmente a su capacidad para capturar relaciones no lineales e interacciones complejas entre las variables de entrada, ya que el modelo se adapta directamente a los datos medidos, sin depender de perfiles de consumo o comportamientos predefinidos.

Si bien *Random Forest* presenta una mayor demanda computacional, asociada principalmente a la etapa de entrenamiento, en el contexto del presente estudio dicha demanda no representa una limitación significativa, dado el tamaño del *dataset* y el horizonte de predicción considerado. No obstante, se deja constancia de que, en escenarios de mayor complejidad, con volúmenes de datos más extensos o requerimientos de entrenamiento frecuente, este aspecto podría convertirse en un factor relevante a tener en cuenta.

Si bien el modelo XGBoost también presentó un desempeño competitivo, con métricas muy cercanas a las obtenidas por *Random Forest*, este último mostró una ligera ventaja en términos de error absoluto (RMSE y MAE) y un mayor coeficiente de determinación (R^2), lo que indica una mejor capacidad para explicar la variabilidad del consumo. Por esta razón, y considerando un desempeño más estable de forma global, *Random Forest* fue seleccionado como el modelo más adecuado para la predicción del consumo energético en el contexto de este estudio.

Por el contrario, el modelo de simulación energética o de ingeniería en DesignBuilder, aunque fue correctamente calibrado a nivel mensual, presentó dificultades para representar el comportamiento real del consumo a escala temporal fina. En especial, no logró representar adecuadamente los picos de demanda asociados al uso puntual de equipos impredecibles y a los hábitos reales de los ocupantes, lo que se tradujo en errores elevados al comparar con los datos medidos.

El análisis del error punto a punto puso en evidencia una alta variabilidad en los dos modelos, con picos significativos a lo largo de día. Esta situación se explica principalmente por la limitada cantidad de datos disponibles para el entrenamiento y por la naturaleza abrupta del consumo residencial, donde pequeños cambios en la actividad de los usuarios puede generar variaciones importantes en la demanda. Aun así, el modelo de *Random Forest* presentó los menores errores relativos, mientras que el modelo de ingeniería mostró peor ajuste frente al consumo real.

Por otra parte, es importante considerar las entradas y requerimientos de información asociados a cada tipo de modelo. En el caso del modelo basado en ingeniería, es decir,

DesignBuilder, se requiere un alto nivel de detalle para la construcción del modelo geométrico, ya que es necesario disponer de las especificaciones arquitectónicas no solo del apartamento, sino también del piso en el que se encuentra, así como de los materiales empleados y los espesores de cada uno de ellos.

En contraste, los modelos basados en datos requieren menos esfuerzo en la etapa de modelado físico y no requieren un nivel de detalle arquitectónico comparable. No obstante, la obtención y preparación del *dataset* implica un proceso que puede ser considerablemente largo y demandante en términos de tiempo. Adicionalmente, como se evidenció en este trabajo, algunos modelos, en particular las redes neuronales, requieren volúmenes de datos mayores para alcanzar un desempeño óptimo, lo cual constituye un aspecto relevante a considerar al seleccionar el enfoque de modelado.

Finalmente, los resultados del estudio confirman que los modelos basados en datos constituyen una herramienta prometedora para predicción del consumo energético en viviendas, especialmente cuando se dispone de información detallada sobre el comportamiento del consumo energético de los usuarios. No obstante, también se evidencia la necesidad de contar con series de datos y medición más extensas y representativas para mejorar la estabilidad de los modelos y reducir la dispersión del error. En este sentido, los enfoques basados en datos y los modelos de ingeniería no deben considerarse excluyentes sino complementarios, ya que la integración puede aportar una visión más completa y realista del desempeño energético de las edificaciones.

Referencias Bibliográficas

- Ahmad, T., Chen, H., & Guo, Y. (2018). Electricity consumption prediction using XGBoost based on discrete wavelet transform. *Proceedings of the International Conference on Advanced Computing and Communication Systems (ICACCS)*.
- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192–1205.
- Barzola-Monteses, J., Parrales-Bravo, C., et al. (2021). Energy consumption of a building using the XGBoost model. *Proceedings of the International Conference on Energy and Environment*.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cabeza, L. F., Bai, Q., Bertoldi, P., Kihila, J. M., Lucena, A. F. P., Mata, É., Mirasgedis, S., Novikova, A., & Saheb, Y. (2022). Buildings. En P. R. Shukla et al. (Eds.), *Climate Change 2022: Mitigation of Climate Change*. Cambridge University Press.
- Chen, C., Gao, Z., Zhou, X., Wang, M., & Yan, J. (2024). Energy consumption prediction and energy-saving suggestions of public buildings based on machine learning. *Energy and Buildings*, 320, 114585.
- 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.
- Gorzałczany, M. B., & Rudziński, F. (2024). Energy consumption prediction in residential buildings: An accurate and interpretable machine learning approach combining fuzzy systems with evolutionary optimization. *Energies*, 17(13), 3242.
- Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2), 357–363.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1), 44–55.
- Iqbal, M. N., Kütt, L., & Rosin, A. (2018). Complexities associated with modeling of residential electricity consumption. *2018 IEEE 59th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*.
- Ji, J., Yu, H., Wang, X., & Xu, X. (2025). Machine learning application in building energy consumption

- prediction: A comprehensive review. *Journal of Building Engineering*, 104, 112295.
- Kowshik, S. B., Ramakrishnan, R., Amutha, S., Dhanush, R., Akshay, P., & Kalyaan, P. P. (2024). Data-driven approach for analysing the relationship between weather patterns and energy consumption in residential buildings for improved energy management. 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS).
- Lahouar, A., & Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, 103, 1040–1051.
- Li, J. (2021). Evolutionary modeling of factors affecting energy consumption of urban residential buildings and prediction models. 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA).
- Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. *Applied Energy*, 162, 1273–1285.
- Ortega-Díaz, L., Cárdenas-Rangel, J., & Osma-Pinto, G. (2023). Estrategias de predicción de consumo energético en edificaciones: Una revisión. *TecnoLógicas*, 26(58), 1–34.
- Revista UIS Ingenierías. (s. f.). Herramienta metodológica para la evaluación energética mediante simulación de edificaciones en el trópico. Portal de Revistas UIS.
- Yesilyurt, H., Dokuz, Y., & Dokuz, A. S. (2024). Data-driven energy consumption prediction of a university office building using machine learning algorithms. *Energy*, 310, 133242.
- Zahedi, R., Aslani, A., Gitifar, A., Farahani, O. N., & Yousefi, H. (2023). Application of artificial neural network in predicting building energy consumption. 2023 8th International Conference on Technology and Energy Management (ICTEM).