

Caracterización y predicción de la demanda energética a corto plazo usando técnicas de  
aprendizaje máquina aplicado al MC - Santander

Juan Camilo Sarmiento Gómez y Sergio Alejandro Uribe Gómez

Trabajo de grado para optar al título de Ingenieros Electrónicos

Director

Juan Manuel Rey López

Doctor en Ingeniería Electrónica

Codirector

Juan Antonio Dominguez Jiménez

Magíster en Ingeniería Eléctrica

Codirector

Cesar Antonio Duarte Gualdrón

Doctor en Ingeniería Eléctrica y Computación

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Ingeniería Electrónica

Bucaramanga

2023

### **Dedicatoria**

La culminación de este proyecto marca el punto más alto de mi viaje para convertirme en ingeniero electrónico, representando el logro más significativo en mi vida hasta ahora. Quiero dedicar esta victoria a dos pilares fundamentales en mi camino: primero, a Dios, quien ha guiado mis pasos y me ha dado fortaleza en cada etapa; y segundo, a mi madre, mi eterna fuente de amor y apoyo incondicional. A lo largo de los desafíos y triunfos, sus enseñanzas han moldeado mi carácter y esencia como persona, brindándome la resiliencia necesaria para enfrentar todas las adversidades que se han cruzado en mi camino.

**Juan Camilo Sarmiento Gómez**

Esta entrega final representa aproximadamente cinco años (o un poco más) de crecimiento personal y académico. Por lo tanto, quiero dedicar este trabajo a las personas que me acompañaron y apoyaron a lo largo de este tiempo. En especial, quiero expresar mi profundo agradecimiento a mis padres, a mi hermano, a mis amigos y familiares cercanos, así como a mi querida mascota. Sé que este logro los llena de orgullo tanto como a mí, y sé que la ilusión siempre estuvo presente. Esto es tanto por mí como por ustedes.

**Sergio Alejandro Uribe Gómez**

### **Agradecimientos**

Quiero expresar mi profundo agradecimiento a Dios y a todas las personas que han sido parte fundamental de mi trayecto académico y personal. En primer lugar, a mi madre, cuyo apoyo inquebrantable me ha sostenido en los momentos de celebración y angustia a lo largo de esta carrera. A Bella, mi amada mascota, que me ha enseñado la paciencia y el amor incondicional, recordándome la importancia de disfrutar de las pequeñas alegrías de la vida.

A todos los amigos que he tenido el privilegio de conocer durante este viaje, les agradezco por su apoyo constante, su amistad sincera y los invaluable momentos compartidos que han enriquecido mi vida.

Finalmente, no puedo pasar por alto mi querida institución, la Universidad Industrial de Santander (UIS), mi alma mater, que ha sido el escenario en el que mi crecimiento no solo como profesional, sino como individuo, se ha acelerado. Las instalaciones de la UIS han sido testigos de momentos esenciales que han forjado mi ser hasta el día de hoy. Gracias a todos ustedes, mi camino ha sido enriquecido de formas innumerables, y este logro es el resultado de su apoyo y guía invaluable.

**Juan Camilo Sarmiento Gómez**

Durante mi tiempo en la carrera, he tenido la fortuna de conocer a personas excepcionales que, de una forma u otra, se han convertido en un grupo en el que siempre he podido confiar. Quiero expresar mi profundo agradecimiento a mis padres, a mi hermano, Sandra, Omar, Andrés,

y familiares cercanos (Amparito y "Taita") por su apoyo constante a lo largo de este viaje. Además, quiero reconocer y agradecer a esos compañeros inquebrantables que estuvieron a mi lado en momentos de adversidad: Brayan, Maria, Natalia, Edgar, Juan Camilo, Jeison, Victor, Oscar, Valentina, Joaquín. Su amistad y apoyo incondicional han sido fundamentales. Este pregrado ha sido un hermoso proceso de crecimiento compartido con ustedes, y espero sinceramente que continuemos este proceso de crecimiento. Gracias por tanto.

**Sergio Alejandro Uribe Gómez**

## Tabla de Contenido

<b>1</b>	<b>Objetivos</b>	<b>14</b>
1.1	Objetivo General	14
1.2	Objetivos Específicos	14
<b>2</b>	<b>Predicción de demanda energética en Santander</b>	<b>15</b>
2.1	Antecedentes locales	15
2.2	Tipos de predicción de demanda energética	16
2.3	Conjunto de datos inicial	18
<b>3</b>	<b>Metodología</b>	<b>20</b>
3.1	Preprocesamiento de los datos	21
3.2	Extracción de características	22
3.2.1	Descomposición empírica en modos (EMD)	22
3.2.2	Datos de tipo calendario	24
3.3	Selección de predictores	27
3.4	Modelos de aprendizaje de máquina	31
3.4.1	Refuerzo Extremo de Gradiente (XGBoost, XGB)	32
3.4.2	Bosque Aleatorio (Random Forest, RF)	34
3.5	Métricas de evaluación	36

<b>4</b>	<b>Análisis de resultados</b>	<b>39</b>
4.1	Predicción con datos reales (Caso 1)	40
4.2	Predicción con datos pronosticados (Caso 2)	43
4.3	Mejores escenarios: Comparativa respecto al Operador de Red (OR) y cálculo de desviaciones	50
4.3.1	Análisis estadístico para modelos destacados	51
4.3.2	Metodología alternativa de ventana ampliada	54
4.3.3	Desviaciones según Acuerdo CNO 1303 de 2020	58
4.3.4	Desviaciones según Resolución CREG 100 de 2019	60
<b>5</b>	<b>Conclusiones y recomendaciones a futuro</b>	<b>62</b>
	<b>Referencias Bibliográficas</b>	<b>65</b>
	<b>Apéndices</b>	<b>71</b>

**Lista de Figuras**

Figura 1	Diagrama de bloques descriptivo del proceso	21
Figura 2	Descomposición en modos empíricos de la demanda energética	26
Figura 3	Diagrama descriptivo del proceso del modelo XGBoost	35
Figura 4	Diagrama descriptivo del proceso del modelo Random Forest	36
Figura 5	Gráfica en el tiempo comparativa entre lo pronosticado por el modelo XGBoost y la demanda real para los meses julio y agosto del año 2023 (Caso 1).	41
Figura 6	Gráfica en el tiempo comparativa entre lo pronosticado por el modelo RF y la demanda real para los meses julio y agosto del año 2023 (Caso 1).	42
Figura 7	Mapa de calor, métrica MAPE diaria para comparativa entre el Operador de Red (OR) y los modelos propuestos en la predicción con datos reales (Caso 1)	44
Figura 8	Gráfica en el tiempo comparativa entre lo pronosticado por el modelo XGBoost y la demanda real para los meses julio y agosto del año 2023 (Caso 2).	47
Figura 9	Gráfica en el tiempo comparativa entre lo pronosticado por el modelo RF y la demanda real para los meses julio y agosto del año 2023 (Caso 2).	48
Figura 10	Mapa de calor, métrica MAPE diaria para comparativa entre el Operador de Red (OR) y los modelos propuestos en la predicción con datos pronosticados (Caso 2)	49
Figura 11	Distribución normal del error para los modelos destacados y operador de red (OR)	52

Figura 12	Histograma de desviación estándar hora-hora para los modelos destacados y operador de red (OR)	52
Figura 13	Cálculo de error acumulado con los mejores modelos y el operador de red en el tiempo (en escala logarítmica)	54
Figura 14	Histograma de desviación estándar hora-hora para los modelos destacados y operador de red (OR) resultante de metodología de ventana ampliada (Caso 2 - VA)	57
Figura 15	Distribución normal del error para los modelos destacados y operador de red (OR) resultante de metodología de ventana ampliada (Caso 2 - VA)	58
Figura 16	Gráfica de barras para el número de tramos por rangos de desviación según resultados bajo el Acuerdo CNO 1303 de 2020.	59
Figura 17	Gráfica de barras para el número de tramos por rangos de desviación según resultados bajo la Resolución CREG 100 de 2019.	61

**Lista de Tablas**

Tabla 1	Algoritmo detallado paso a paso del método EMD.	25
Tabla 2	Hiperparámetros utilizados para los algoritmos de selección de predictores	28
Tabla 3	Predictores seleccionados por el algoritmo de Eliminación Recursiva de Características (RFE)	30
Tabla 4	Predictores seleccionados por el algoritmo de selección secuencial de características hacia al frente (SFS)	30
Tabla 5	Predictores seleccionados por el algoritmo de selección secuencial de características hacia atrás (SBS)	31
Tabla 6	Recopilación de desempeño para los modelos en Caso 1	45
Tabla 7	Recopilación de desempeño para los modelos en Caso 2	46
Tabla 8	Predictores seleccionados por el algoritmo de selección secuencial de características hacia al frente (SFS) para un horizonte de una semana (Ventana ampliada, Caso 2 - VA)	56
Tabla 9	Recopilación de desempeño para los mejores modelos para Caso 2, Caso 2 - VA y OR	57
Tabla 10	Tabla recopilatoria del número de tramos por rangos de desviación según resultados bajo el Acuerdo CNO 1303 de 2020.	60

Tabla 11 Tabla recopilatoria del número de tramos por rangos de desviación según resultados bajo la Resolución CREG 100 de 2019.

61

**Lista de Apéndices**

	<b>pág.</b>
Apéndice A    Repositorio GitLab CPDEAM222	71
Apéndice B    Desviaciones calculadas bajo el Acuerdo CNO 1303 de 2020 con los resultados de los modelos XGBoost-SFS, RF-SFS y el Operador de Red (OR)	72
Apéndice C    Desviaciones calculadas bajo la Resolución CREG 100 de 2019 con los resultados de los modelos XGBoost-SFS, RF-SFS y el Operador de Red (OR)	75

## Resumen

**Título:** Caracterización y predicción de la demanda energética a corto plazo usando técnicas de aprendizaje máquina aplicado al MC - Santander. \*

**Autores:** Juan Camilo Sarmiento Gómez, Sergio Alejandro Uribe Gómez. \*\*

**Director:** Juan Manuel Rey López. Dr. Ingeniería Electrónica.

**Co-Directores:** Juan Antonio Dominguez Jiménez, MsC. Ingeniería Eléctrica. Cesar Antonio Duarte Gualdrón, Dr. Ingeniería Eléctrica y Computación.

**Palabras Clave:** Predicción, Demanda, Santander, Aprendizaje de máquina, STLF.

**Descripción:** La energía eléctrica es un recurso indispensable en el día a día de la sociedad, pues prácticamente la totalidad de los procesos cotidianos, a nivel residencial e industrial dependen de esta. Por esta razón, es necesario un sistema de distribución robusto y confiable, con una alta capacidad de planeación en la generación, despacho y transporte de la energía eléctrica, cuyo pilar fundamental es un correcto pronóstico en la demanda energética. En este trabajo de grado se propone el desarrollo de un modelo predictivo de tres fases para la predicción de demanda energética en el departamento de Santander: extracción de características de las entradas de la máquina, selección de predictores, y predicción, donde se optó por métodos de aprendizaje automático por conjuntos basados en árboles de decisión. Finalmente, el desempeño de los modelos se midió bajo las métricas MAPE y RMSE.

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Fisicomecánicas, Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director Juan Manuel Rey Lopez. Codirector Juan Antonio Dominguez Jiménez. Codirector Cesar Antonio Duarte Gualdrón.

## Abstract

**Title:** Characterization and Short-Term Energy Demand Forecasting Using Machine Learning Techniques Applied to MC - Santander. \*

**Authors:** Juan Camilo Sarmiento Gómez, Sergio Alejandro Uribe Gómez. \*\*

**Director:** Juan Manuel Rey Lopez, Ph.D in Electrical Engineering.

**Co-directors:** Juan Antonio Dominguez Jiménez, MsC in Electrical Engineering. Cesar Antonio Duarte Gualdrón, Ph.D in Electronical Engineering and .

**Keywords:** Forecasting, Demand, Santander, Machine Learning, STLF.

**Description:** Electricity is an indispensable resource in the daily life of society, as practically all daily, both at the residential and industrial levels, depend on it. For this reason, a robust and reliable distribution system is necessary, with a high capacity for planning in electricity generation, dispatch, and transportation, whose fundamental pillar is accurate energy demand forecasting. In this bachelor thesis is proposed the development of a three-phase predictive model for energy demand prediction in the Santander department: feature extraction, predictor selection, and prediction, where ensemble machine learning methods based on decision trees were chosen. Finally, the performance was measured using the RMSE and MAPE criteria.

---

\* Bachelor Thesis

\*\* Faculty of Physical and Mechanical Engineering, School of Electrical, Electronic and Telecommunications Engineering. Director Juan Manuel Rey López. Codirector Juan Antonio Dominguez Jiménez. Codirector Cesar Antonio Duarte Gualdrón.

## **1. Objetivos**

### **1.1. Objetivo General**

Diseñar un modelo predictivo que permita la caracterización y pronóstico de la demanda de energía eléctrica en el MC - Santander utilizando técnicas de aprendizaje de máquina.

### **1.2. Objetivos Específicos**

- Implementar un algoritmo encargado del proceso de caracterización de las entradas de la máquina, comprendiendo el preprocesamiento de los datos de entrada y la selección de predictores del sistema.
- Implementar un algoritmo de pronóstico de demanda de energía eléctrica utilizando técnicas de aprendizaje de máquina.
- Evaluar el desempeño del modelo predictivo conformado por el conjunto de algoritmos implementados utilizando como datos de prueba los ofrecidos por XM S.A ESP para el mercado de comercialización (MC) de Santander.

## **2. Predicción de demanda energética en Santander**

### **2.1. Antecedentes locales**

La predicción de demanda energética es un tema que involucra a los sistemas y mercados eléctricos nacionales e internacionales, al ser un aspecto crítico y de gran relevancia para modelos de planificación de generación y suministro de electricidad (Taylor et al., 2006). Por esta razón, reducir el margen de error al realizar predicciones se ha convertido en un tópico muy importante, ya que una estimación incorrecta puede conllevar a que la entidad encargada del suministro energético (en Colombia, el operador de red, OR) sea penalizada. Esto ha llevado a la reciente exploración de soluciones que permitan minimizar de manera eficaz esta problemática. Dichos métodos se encuentran frecuentemente encabezados por el uso de inteligencia artificial dada su creciente popularidad al lograr reducir el error de estimación frente a las técnicas convencionales (Zor et al., 2017), conocidas como técnicas analíticas.

Frente al contexto nacional colombiano, el departamento de Santander ha sido uno de los focos principales de atención para realizar estudios en este tópico, encabezados principalmente por la Universidad Industrial de Santander (UIS); entre los diversos estudios realizados se destaca la existencia de una guía para la predicción de demanda energética en el departamento (Galán and Díaz, 2021) donde se sugieren pautas para el tratamiento de datos, y se identifican las variables a considerar al momento de realizar predicción de demanda a corto plazo en el departamento. También se expone la necesidad urgente de este tipo de estudios debido a que, aparte de reducir el margen de error para evitar penalizaciones, el operador de red (OR), en este caso la ESSA,

contaba para el año 2018 con el uso de métodos empíricos para realizar la predicción (Galán and Díaz, 2021). Entre las diferentes técnicas propuestas desde la UIS para la solución de esta problemática utilizando métodos de inteligencia artificial, se encuentra el uso de redes neuronales artificiales (Arenas, 2004) (Agudelo and Barón, 2022), y las máquina de soporte vectorial (Perez and Tarazona, 2021). La proposición de diversas técnicas, en conjunto al tiempo de estudio del tópico sugiere reconocer a la entidad (UIS) como una de las principales fuente de información para este tipo de proyecto, y por lo tanto, uno de los pilares fundamentales para la ejecución de este proyecto. Proyectos anteriores también contemplan el Acuerdo CNO 1303 de 2020 y la Resolución CREG-100 de 2019 para evaluar el comportamiento de los modelos propuesto por lo que resulta imperativo considerar estas regulaciones.

## **2.2. Tipos de predicción de demanda energética**

Es importante identificar que de acuerdo al horizonte en el que se planea realizar el pronóstico, se puede clasificar la predicción de demanda energética en cuatro grupos principales: largo plazo, mediano plazo, corto plazo y muy corto plazo; sin embargo, depende de la literatura consultada y las legislaciones locales el escoger cuál se ajusta más al caso de estudio vigente. A continuación se presentan brevemente algunas definiciones acerca de los tipos de predicción de demanda energética anteriormente mencionados:

- Pronóstico de demanda a largo plazo: Predicción en tiempos mayores a un año; puede ser de un año hasta décadas (Villamil, 2016), de tres años a décadas (Zor et al., 2017) o de cinco años (Cardona, 2021).

- Pronóstico de demanda a mediano plazo: Predicción con horizonte de dos semanas a tres años (Zor et al., 2017) o de varios meses hasta un año (Villamil, 2016). Usualmente en Colombia corresponde a la predicción que se realiza para el despacho de cinco semanas (Cardona, 2021).
- Pronóstico de demanda a corto plazo (\*): Conocido como Short-Term Load Forecasting (o por sus siglas, STLF) en el idioma inglés se refiere a la predicción con horizonte horaria, diaria o semanal (Zor et al., 2017), de una semana o dos semanas (Villamil, 2016) o sencillamente en un plazo de veinticuatro horas (Cardona, 2021). Este tipo de predicción es la temática principal de este trabajo de investigación, donde se acoge la predicción en un plazo de veinticuatro horas bajo un ejercicio que se presentará más adelante llamado 'Caso 1' y semanal en 'Caso 2' con el fin de asemejarse en lo máximo a cómo realizan el despacho energético en Colombia los operadores de red (OR) y según lo regulado por la Resolución CREG-025 de 1995.
- Pronóstico de demanda a muy corto plazo: Puede referirse al pronóstico realizado de minutos hasta una hora por máximo (Zor et al., 2017) o desde la hora actual del día hasta el fin del mismo (Cardona, 2021).

---

El uso del símbolo (\*) que se usa en conjunto a la definición de pronóstico de demanda a corto plazo es utilizado con el fin de denotar el concepto de mayor interés en esta sección.

### 2.3. Conjunto de datos inicial

El conjunto de datos utilizado para el desarrollo del modelo de pronóstico en el inicio contaba únicamente con los datos de demanda energética real hora-hora del departamento de Santander (los cuales se pueden encontrar como información pública en los reportes mensuales que publica el operador de red en el portal de XM S.A ESP) en conjunto a la fecha y tipo de día. Esta última es una variable que se utiliza recurrentemente en Colombia con el fin de marcar las tendencias entre los días ordinarios, fines de semana y festivos tal como es explicado en Galán and Díaz (2021). También fueron incluidas variables meteorológicas debido a que son de uso recurrente y puede llegar a correlacionarse en gran medida con la demanda energética como sugieren estudios nacionales e internacionales (Villamil, 2016). Las variables meteorológicas contempladas son la temperatura promedio, la humedad y la presión registradas con frecuencia horaria para las veinticuatro horas del día. Cabe aclarar que la temperatura, la humedad y la presión corresponden a la de la ciudad de Bucaramanga exclusivamente debido a que la demanda de energía en el departamento de Santander está concentrada en gran parte en el área metropolitana de Bucaramanga con un sesenta por ciento (60%), y ha demostrado ser eficiente para el cálculo de proyección en Santander (Galán and Díaz, 2021); aunque en trabajos anteriores la temperatura ha sido tomada desde el Aeropuerto de Palonegro, Lebrija como en Agudelo and Barón (2022), en este trabajo se optó por utilizar variables meteorológicas censadas directamente desde Bucaramanga, exactamente en el edificio de Eléctrica, Electrónica y Telecomunicaciones (E3T) de la Universidad Industrial de Santander, debido a que al tratar de consultar esta información en portales disponibles en la red se encontró

con que estos datos rondaban desde las 6:00 A.M y las 11:00 P.M , quedando un espacio de información en incertidumbre desde la medianoche. En contraste, la información censada en la E3T sí correspondía a las veinticuatro horas del día.

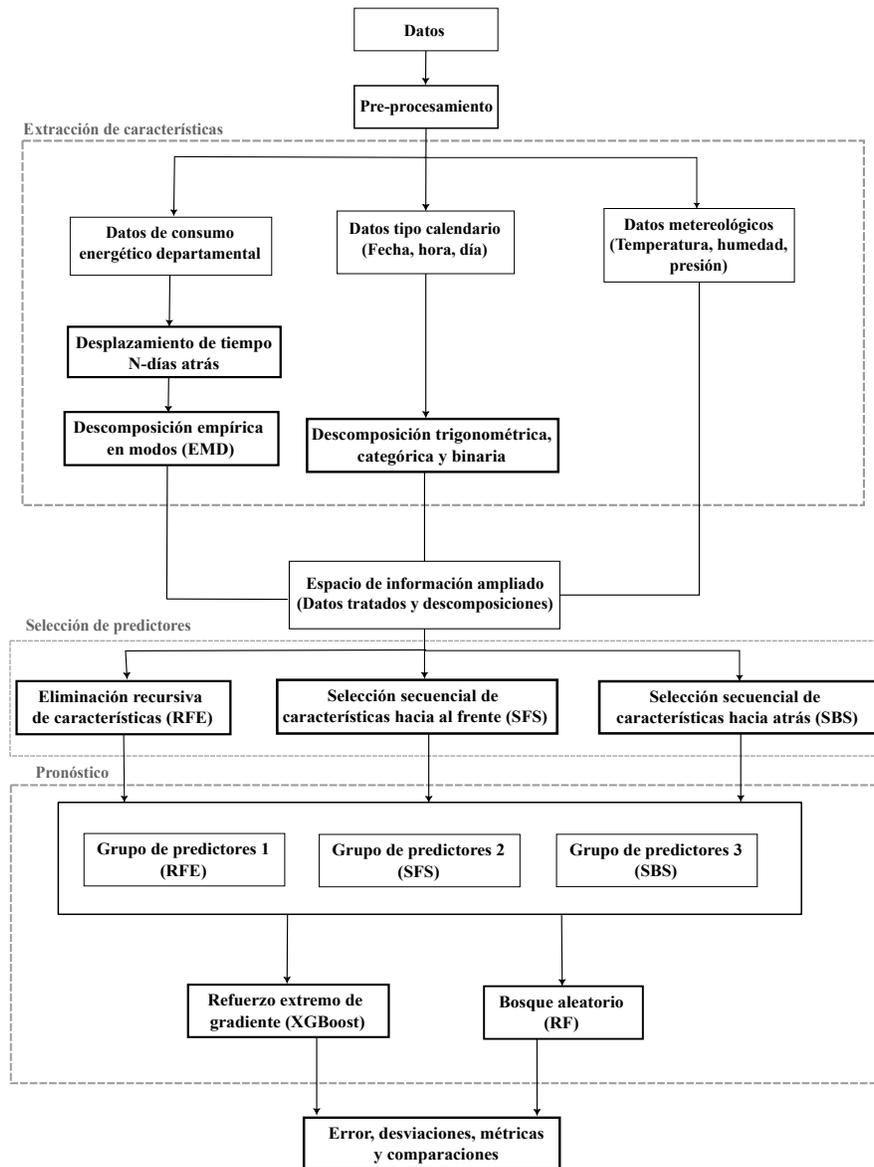
El conjunto de datos inicial contiene los datos meteorológicos, de demanda, de calendario, horario y tipo de día desde 1 de enero de 2023 hasta el día 31 de agosto de 2023, para un total de ocho (8) meses.

### 3. Metodología

La metodología propuesta para llevar a cabo este proyecto consta del uso de un modelo de varias etapas en cascada, también llamadas híbridas o adaptativas debido a que el uso de este tipo de metodologías con etapas enfocadas en clasificación y agrupamiento de datos previamente a la predicción suponen un mejoramiento de los pronósticos de demanda energética como ha sido demostrado en contextos nacionales (Cardona, 2021) e internacionales (Tajeuna et al., 2018). La metodología propuesta consta de tres etapas principales: extracción de características (o caracterización) que tiene como propósito aumentar el espacio de información inicial, mediante la extracción de características relevantes en el conjunto de datos de entrada; selección de predictores es una etapa que busca escoger las características más importantes que pueden encontrarse en el conjunto de datos; y, finalmente la etapa de pronóstico que es donde se busca, a través de modelos de aprendizaje de máquina que utilizan técnicas de regresión, realizar la predicción utilizando los predictores más relevantes seleccionados en la etapa anterior. El modelo de tres etapas se ve precedido de una fase inicial llamada preprocesamiento y, luego de haber sido realizado el proceso anteriormente descrito se realiza una etapa de evaluación bajo diversas métricas; esta información descrita puede ser visualizada en el diagrama de bloques expuesto en la Figura 1.

La implementación del modelo fue realizada en su totalidad utilizando el lenguaje de programación Python para cada etapa (Ver Apéndice A del presente documento).

**Figura 1**  
*Diagrama de bloques descriptivo del proceso*



### 3.1. Preprocesamiento de los datos

Como fue expuesto anteriormente, la lista de predictores candidatos para la máquina está compuesta por datos meteorológicos tomados del edificio de ingeniería eléctrica de la Universi-

dad Industrial de Santander, y datos característicos de la demanda energética del departamento de Santander publicados en la página de XM S.A ESP, empresa encargada de la administración y operación del mercado energético colombiano. El formato de publicación de los datos son archivos de extensión .xlsx para datos meteorológicos y de demanda energética. El formato y la organización de los datos hizo necesario la reorganización y el preprocesamiento de estos para espaciarlos de manera uniforme y eliminar posible valores no numéricos (NaN).

Para los datos correspondientes a la demanda energética del departamento, se realizó un algoritmo en Python encargado de extraerlos automáticamente de los archivos .xlsx, almacenando la demanda registrada con frecuencia horaria, su fecha de registro y el tipo de día correspondiente a la fecha.

Para los datos meteorológicos, se realizó un algoritmo en Python encargado de extraerlos automáticamente de los archivos .xlsx, almacenando la temperatura, la humedad y la presión registradas con una frecuencia de 10 minutos. La frecuencia de registro de los datos hizo necesario calcular la media horaria para cada predictor meteorológico. En el caso de las horas donde no se tenía registro de datos (valores NaN) fueron reemplazados por la media de la variable registrada.

Por último, teniendo los datos organizados como series temporales, se almacenaron en un dataframe, estructura de datos bidimensional ampliamente utilizada en el análisis de datos, teniendo una primera versión del espacio de información de posibles predictores para la máquina.

## **3.2. Extracción de características**

**3.2.1. Descomposición empírica en modos (EMD).** Un enfoque para mejorar el rendimiento de los métodos de aprendizaje de máquina es la extracción de características, que

consiste en extraer la información más relevante de las entradas de la máquina para describir un gran conjunto de datos (Apolinario et al., 2021). En el campo de la predicción energética, las tareas de pronóstico suelen tener una estructura similar, donde es destacable la importancia de las técnicas utilizadas para la extracción de características de los datos históricos de consumo energético (Zhang et al., 2017).

La descomposición empírica en modos (EMD) es uno de los métodos más utilizado en la extracción de características (Apolinario et al., 2021). En Bican and Yaslan (2014), Yu et al. (2008) y Fan and Tang (2013), el EMD ha sido utilizado para pronóstico de series de tiempo.

Teniendo en cuenta la revisión bibliográfica previa a la implementación del modelo de predicción de demanda, se escogió el EMD como método principal de extracción de características de los datos históricos de demanda energética en el departamento de Santander.

El EMD es un método de análisis espacio-temporal adaptativo adecuado para procesar datos de series temporales que no son estacionarios ni lineales, dividiendo las series de tiempo en modos en el dominio temporal (Apolinario et al., 2021). Básicamente, descompone una señal en varias series de tiempo, brindando la posibilidad de modelar cualquier señal compleja a través de su conjunto finito de funciones de modo intrínseco (IMF). Esto permite extraer las características más influyentes en el patrón de comportamiento de la demanda energética en Santander.

El objetivo del método EMD es descomponer los datos de series de tiempo no-lineales y no-estacionarias en varias componentes Bican and Yaslan (2014) como en la ecuación 1.

$$x(t) = \sum_{i=1}^n h_i + r \quad (1)$$

Donde  $x(t)$  es el valor original del histórico de datos,  $h_i$  representa cada IMF y  $r$  es la componente residual Bican and Yaslan (2014). En Huang et al. (1998) se puede encontrar una explicación más detallada sobre el EMD, y la matemática y teoría de señales detrás del método.

Para el proceso de extracción de características se utilizó la información de 8 IMFs para nutrir el espacio de información con los predictores candidatos para las máquinas. La Figura 2 muestra las IMFs de los datos históricos de demanda utilizados para el entrenamiento de las máquinas. Se puede observar que la información contenida en las IMFs pasa de características de alta a baja frecuencia.

Teniendo en cuenta el desglose de la algoritmia detrás del EMD en Bican and Yaslan (2014) y Wu et al. (2016), así como revisión bibliográfica previa, se detalló el paso a paso del método EMD en la Tabla 1. El modulo de Python utilizado en el presente trabajo corresponde a EMD-signal Laszuk (2017).

**3.2.2. Datos de tipo calendario.** Las variables como los días de trabajo, fines de semana y festivos pueden cambiar notablemente el patrón de consumo de energía, Moral-Carcedo and Pérez-García (2019) y Ziel (2018). Por esta razón, y por el patrón cíclico temporal observado en el consumo de demanda energética se tuvieron en cuenta las variables calendario del conjunto de datos. Se tuvieron variables categóricas para el día de la semana y las horas del día. También se utilizó la variable tipo de día, de la cual se tomó en cuenta únicamente los días festivos de la

**Tabla 1**

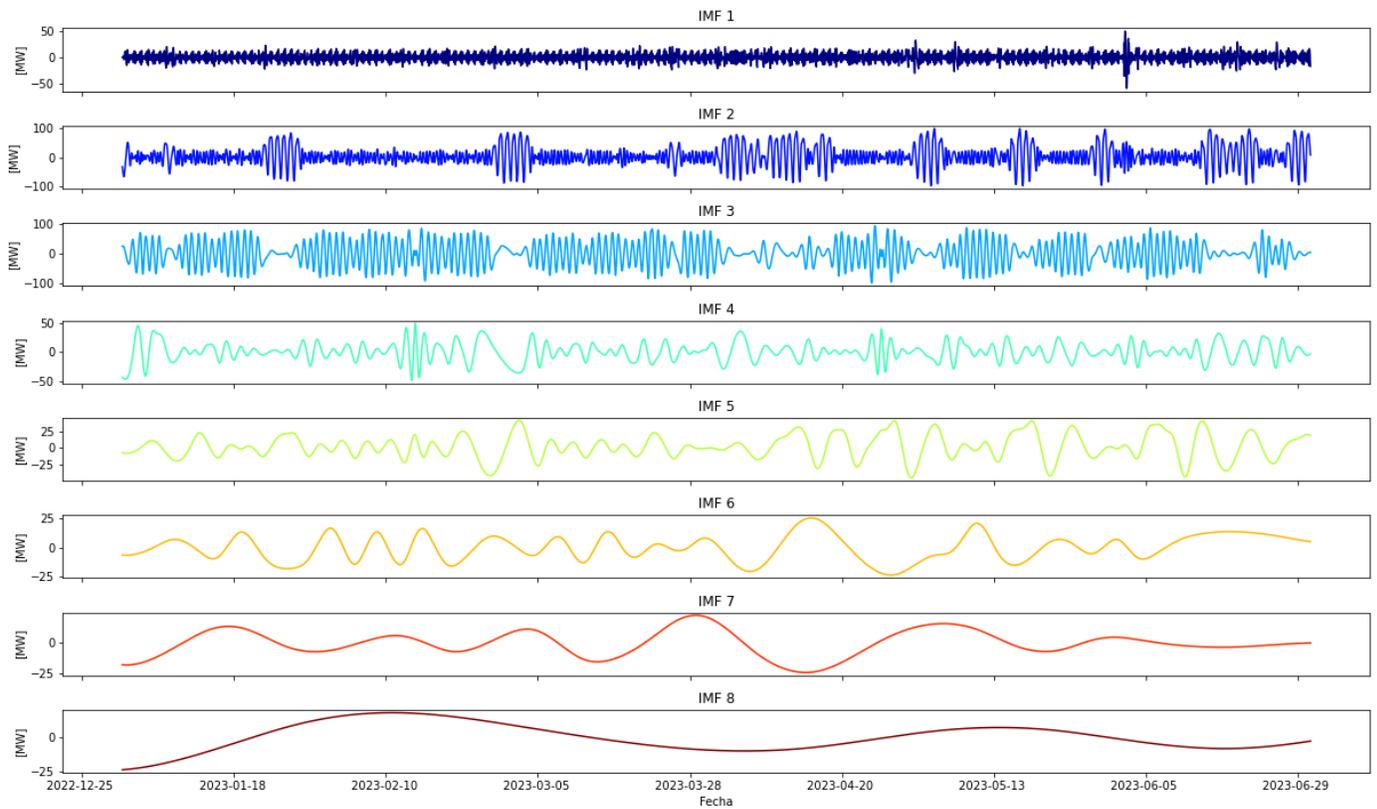
*Algoritmo detallado paso a paso del método EMD.*

<b>Algoritmo EMD</b>
<b>Paso 1:</b> Calcular las muestras de mínimos locales $x_l[n]$ y máximos locales $x_u[n]$ de la señal de entrada.
<b>Paso 2:</b> Interpolar $x_l[n]$ y $x_u[n]$ para construir las funciones envolventes inferior y superior, $x_l(t)$ y $x_u(t)$ respectivamente.
<b>Paso 3:</b> Calcular la media de las funciones envolventes. Por ejemplo, $m_1(t) = [x_l(t) + x_u(t)]/2$ .
<b>Paso 4:</b> Calcular la diferencia, $h_1(t) = x(t) - m_1(t)$ .
<b>Paso 5:</b> Verificar si $h_1(t)$ satisface las dos condiciones de la propiedad de una IMF Huang et al. (1998). 1) El número de extremos y cruces por cero debe ser igual o diferir como máximo en uno. 2) En cualquier punto, el valor medio definido por los máximos y mínimos locales debe ser cero.
<b>Paso 6:</b> Si no se satisfacen las condiciones del paso 5, se repiten los pasos del 2 al 5 hasta que se encuentre una función IMF válida $h_k(t)$ .
<b>Paso 7:</b> Calcular el primer residuo $res_1(t) = x_1(t) - h_k(t)$ .
<b>Paso 8:</b> Usar $res_1(t)$ como nuevos datos de series temporales para calcular las siguientes IMFs hasta adquirir un residuo creciente monótono $res_n(t)$ .

*Nota.* Esta tabla describe paso a paso el funcionamiento del algoritmo EMD. Adaptado de Bican and Yaslan (2014).

semana.

Para el caso de las variables categóricas correspondientes a la hora del día, y el día de la semana, se utilizaron dos aproximaciones. La primera, utilizarlas el formato categórico, en el caso de la hora variando está en orden creciente en el rango de [0-23], y en el caso del día de la semana variando este en orden creciente en el rango de [0-6]. La segunda aproximación fue utilizar regresores periódicos definidos por series de tiempo trigonométricas, aproximación utilizada en Dab et al. (2019), Weron (2007), WH (1921) y Moral-Carcedo and Pérez-García (2017). Incluyendo señales periódicas, se espera que el modelo pueda captar patrones diarios y horarios repetitivos en

**Figura 2***Descomposición en modos empíricos de la demanda energética*

el consumo de demanda energética (Dab et al., 2019). Las variables explicativas trigonométricas están definidas en las Ecuaciones 2 y 3.

$$\gamma_{i,t} = \sin(\omega_i t) \quad (2)$$

$$\phi_{i,t} = \cos(\omega_i t) \quad (3)$$

En las Ecuaciones 2 y 3,  $\omega_i = \frac{2\pi}{T}$ ,  $T = 24$  es el periodo para el tiempo definido en horas,

y  $T = 7$  el periodo para el tiempo definido en días.

Para el caso de los días festivos, se tomaron de la variable tipo de día las fechas de festividad en Colombia, y los fines de semana (sábado y domingo) de la variable categórica día de la semana. Para incluirlos como variables de entrada numéricas para las máquinas, se realizó una serie binaria de tiempo  $X_t$ , donde  $X_t = 1$  para los días clasificados como festivos, y  $X_t = 0$  para el resto de días.

### 3.3. Selección de predictores

Debido a la existencia de un alto campo de información el cual cuenta con 114 características luego de haber sido realizado el aumento de datos utilizando las técnicas de descomposición trigonométrica, empírica en modos y categorización y desplazamiento en el tiempo de hasta seis días atrás, existe una alta complejidad para escoger las características que probablemente sean las más efectivas para que los algoritmos de aprendizaje de máquina absorban las tendencias que existentes en las series de tiempo de demanda energética para la predicción, por lo cual, se optó en utilizar métodos que se encarguen de esta labor. Los algoritmos escogidos fueron Eliminación Recursiva de Características (Recursive Feature Elimination, RFE), Selección Secuencial hacia adelante (Sequential Forward Selection, SFS) y Selección Secuencia hacia atrás (Sequential Backward Selection, SBS) los cuales a través de la iteración recurrente buscan eliminar la característica menos importante en el conjunto de datos bajo el uso de una métrica de rendimiento o coeficiente. Los hiperparámetros establecidos para la ejecución de cada método de selección de predictores se muestran en la Tabla 2.

En la literatura, es notable la escasa presencia de aplicaciones que involucren la combinación de los tres métodos para esta aplicación en particular. No obstante, existen ejemplos en otros

**Tabla 2***Hiperparámetros utilizados para los algoritmos de selección de predictores*

<b>Hiperparámetros utilizados para los algoritmos de selección de predictores</b>				
<b>Método</b>	<b>Hiperparámetro</b>			
	estimator	k_features o n_features	step	scoring
RFE	Regresión Lineal	10	1	No aplica
SFS	Regresión Lineal	10	No aplica	Neg Mean Squared Error
SBS	Regresión Lineal	10	No aplica	Neg Mean Squared Error

campos, como la predicción del cáncer de seno, donde se han empleado con éxito los tres algoritmos simultáneamente y se han obtenido resultados altamente positivos (Sweetlin and Saudia, 2023). De manera individual, el método RFE ha demostrado ser una opción adecuada en la predicción de la demanda energética a corto plazo, especialmente cuando se utiliza en conjunto con un modelo de inteligencia artificial, como se evidencia en Prashanthi and Priyadarsini. (2021).

En las Tablas 3, 4 y 5 se pueden observar los resultados de la ejecución de dichos algoritmos, dichas tablas contienen el nombre de la variable (acrónimo) al lado derecho, y, al lado izquierdo a qué corresponde esta característica; como es notable, se optó por seleccionar únicamente diez características con cada uno de los algoritmos.

Al analizar las tablas de resultados, se destaca que los predictores seleccionados por cada uno de los métodos presentan diferencias notables, a pesar de que la mayoría de ellos están relacionados con la demanda energética. Esto puede explicarse por el hecho de que el espacio de información inicial consta de un gran número de características (114), lo que conlleva la posibilidad de que muchas de estas sean redundantes o poco relevantes. Además, algunas características relevantes pueden estar correlacionadas entre sí, lo que influye en que los diferentes métodos se-

leccionen conjuntos distintos de predictores que representan una misma tendencia.

Es importante tener en cuenta que, a diferencia de Sequential Forward Selection (SFS), que comienza con un conjunto vacío de características e incorpora de forma iterativa aquellas que mejoran el desempeño del modelo, Recursive Feature Elimination (RFE) y Sequential Backward Selection (SBS) inician con un conjunto completo de predictores y proceden a eliminarlos de manera iterativa. En las tablas de resultados, se observa que prácticamente todos los predictores seleccionados por RFE y SBS están relacionados con la demanda energética, ya sea de manera directa o a través de una IMF. Esto sugiere que, teniendo ambos métodos un enfoque similar, tienden a seleccionar características diferentes pero representativas de una misma tendencia.

En contraste, SFS no se limita exclusivamente a las variables relacionadas con la demanda, sino que también incluye datos calendario y meteorológicos en su selección. Esto indica una relevancia en la forma de selección de predictores (incorporación o eliminación) en relación al tipo de predictores seleccionados.

La elección de estos algoritmos para realizar la selección de predictores se debe a que en estudios anteriores han demostrado contar con una alta eficiencia para este tipo de estudios y su uso habitual en estudios internacionales de este tipo. Es de destacar que los tres algoritmos utilizados para esta etapa escogieron funciones de modo intrínseco (IMF) como potenciales características relacionadas directamente con la demanda energética aún teniendo la posibilidad de escoger la demanda directamente, por lo que, probablemente este tipo de funciones sí realicen un aporte significativo.

**Tabla 3***Predictores seleccionados por el algoritmo de Eliminación Recursiva de Características (RFE)*

<b>Predictores escogidos por el algoritmo RFE</b>	
<b>Predictor</b>	<b>Acrónimo</b>
Festivos	holiday
Cuarta función de modo intrínseco para la demanda dos días antes	demand_2_days_before_IMF_4
Sexta función de modo intrínseco para la demanda dos días antes	demand_2_days_before_IMF_6
Séptima función de modo intrínseco para la demanda dos días antes	demand_2_days_before_IMF_7
Demanda dos días antes del día evaluado	demand_2_days_before
Quinta función de modo intrínseco para la demanda tres días antes	demand_3_days_before_IMF_5
Sexta función de modo intrínseco para la demanda tres días antes	demand_3_days_before_IMF_6
Séptima función de modo intrínseco para la demanda tres días antes	demand_3_days_before_IMF_7
Sexta función de modo intrínseco para la demanda cuatro días antes	demand_4_days_before_IMF_6
Demanda cuatro días antes del día evaluado	demand_4_days_before

**Tabla 4***Predictores seleccionados por el algoritmo de selección secuencial de características hacia al frente (SFS)*

<b>Predictores escogidos por el algoritmo SFS</b>	
<b>Predictor</b>	<b>Acrónimo</b>
Hora (En formato categórico, de la hora 0 a la 23 del día)	hour
Festivos	holiday
Temperatura (en grados Celsius)	Temp
Día (En formato categórico, del día 0 al día 6)	t5
Demanda un día antes del día evaluado	demand_1_days_before
Demanda tres días antes del día evaluado	demand_3_days_before
Primera función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_1
Segunda función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_2
Tercera función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_3
Cuarta función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_4

**Tabla 5**

*Predictores seleccionados por el algoritmo de selección secuencial de características hacia atrás (SBS)*

<b>Predictores escogidos por el algoritmo SBS</b>	
<b>Predictor</b>	<b>Acrónimo</b>
Segunda función de modo intrínseco para la demanda un día antes	demand_1_days_before_IMF_2
Tercera función de modo intrínseco para la demanda un día antes	demand_1_days_before_IMF_3
Cuarta función de modo intrínseco para la demanda un día antes	demand_1_days_before_IMF_4
Quinta función de modo intrínseco para la demanda un día antes	demand_1_days_before_IMF_5
Demanda un día antes del día evaluado	demand_1_days_before
Segunda función de modo intrínseco para la demanda dos días antes	demand_2_days_before_IMF_2
Tercera función de modo intrínseco para la demanda dos días antes	demand_2_days_before_IMF_3
Demanda dos días antes del día evaluado	demand_2_days_before
Segunda función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_2
Tercera función de modo intrínseco para la demanda seis días antes	demand_6_days_before_IMF_3

### 3.4. Modelos de aprendizaje de máquina

Para la etapa de pronóstico fueron escogidos dos modelos que se basan en el uso de árboles de decisión: Bosque Aleatorio (Random Forest, RF, en inglés) y Refuerzo de Gradiente Extremo (XGBoost, XGB, en inglés). Este tipo de métodos fueron elegidos debido a que en Santander hasta el momento no han sido ampliamente explorados y, a su vez, este tipo de algoritmos han demostrado tener resultados satisfactorios en este tipo de problemas. Para la validación en el entrenamiento de ambos modelos se utilizó la validación cruzada para series de tiempo, la cual tiene en cuenta el carácter temporal de las predicciones. El entrenamiento del modelo se realizó con datos de enero hasta junio de 2023. El mes de julio y agosto de 2023 fue utilizado para la etapa de prueba.

Para la predicción con series de tiempo no resulta conveniente permitir que la validación

cruzada tome datos aleatoriamente dado el enfoque de este proyecto, debido a que podría causar que los modelos aprendan erróneamente del conjunto de datos y haya sesgo.

A continuación, en las siguientes subsecciones es posible encontrar una breve descripción de cada método.

**3.4.1. Refuerzo Extremo de Gradiente (XGBoost, XGB).** El Gradient Boosting, o Refuerzo de Gradiente, es un potente algoritmo de aprendizaje automático por conjuntos utilizado tanto en tareas de clasificación como en regresión. Esta técnica combina los resultados de varios aprendices base para construir un modelo que suele ofrecer un mejor rendimiento que los modelos de aprendizaje automático convencionales. Por lo general, el Refuerzo de Gradiente utiliza árboles de decisión como aprendices base (Paliari et al., 2021).

El XGBoost, que significa Refuerzo de Gradiente Extremo, es una implementación específica del Refuerzo de Gradiente desarrollada inicialmente como proyecto de investigación en la Universidad de Washington (Zheng et al., 2017). Lo que diferencia a XGBoost de las técnicas tradicionales de Refuerzo de Gradiente son los detalles específicos de su implementación, lo que le confiere una mayor potencia y, por ende, un mejor rendimiento. Uno de estos aspectos es la utilización de técnicas avanzadas de regularización, que reducen el sobreajuste y mejoran la generalización del modelo (Paliari et al., 2021).

En términos generales, XGBoost ha demostrado un rendimiento superior en comparación con el Refuerzo de Gradiente convencional, como se evidencia en estudios comparativos directos Kadiyala and Kumar (2018). A pesar de ser una técnica relativamente nueva, publicada en 2015, la literatura existente en el área de pronóstico energético sobre su uso es limitada. Esto presenta una

oportunidad interesante para evaluar el desempeño de este método en el contexto de la predicción energética en el departamento de Santander.

La base teórica de nuestro modelo se fundamenta en las investigaciones previas realizadas por Anggraeni et al. (2021) y Phan et al. (2020). El objetivo principal de XGBoost es iterar de manera continua para mejorar el modelo de árbol en cada paso, reduciendo su función objetivo. XGBoost genera un modelo final utilizando el descenso del gradiente en cada árbol de decisión, lo que resulta en la creación de múltiples modelos, cada uno de los cuales corrige el modelo anterior, como se describe en (Anggraeni et al., 2021). El resultado predicho por XGBoost se obtiene sumando las predicciones de todos los árboles generados, como se ilustra en la Figura 3.

Como se puede apreciar en la Figura 3, XGBoost se caracteriza por ser un modelo aditivo que utiliza un conjunto de  $M$  árboles de decisión. En esta representación,  $f$  representa cada árbol de decisión individual, y  $F$  es la función resultante que combina todas las contribuciones de estos árboles. En el proceso de regresión, la función objetivo del modelo está dada por la ecuación 4.

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m),$$

$$\theta = (f_1, f_2, \dots, f_M) \quad (4)$$

En la ecuación 4,  $l$  representa la función de pérdida, mientras que  $\Omega$  corresponde al término de regularización. La asignación de vectores se utiliza para mejorar el árbol de decisión en relación con cada término de regularización  $\Omega(f)$ . Por lo tanto,  $\Omega(f)$  se define mediante la ecuación 5.

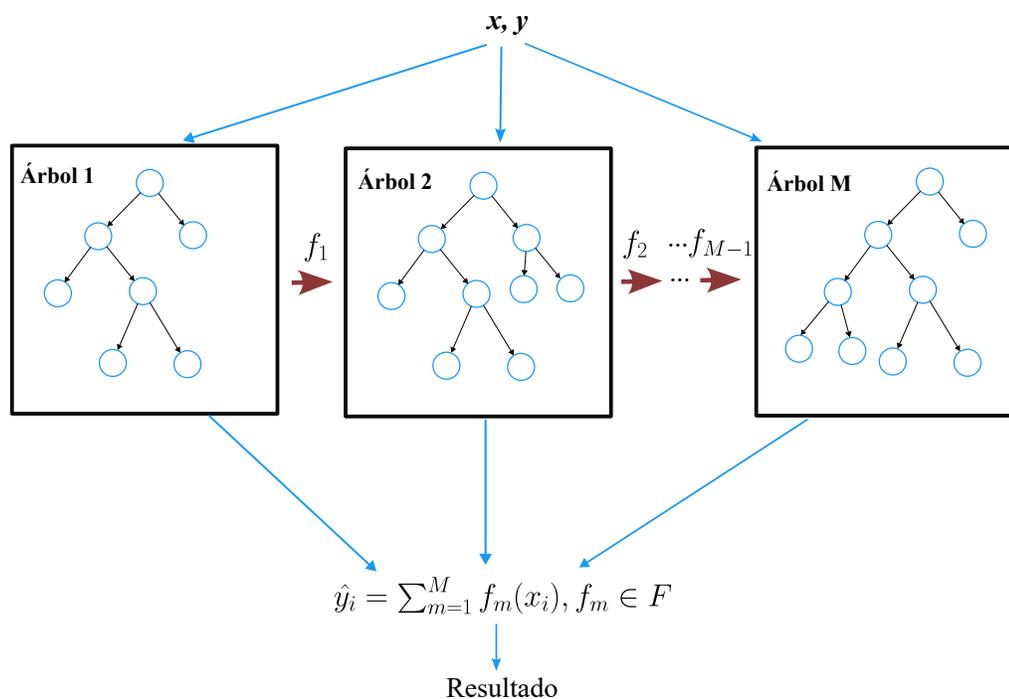
$$\Omega(f) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^T \omega_j^2 \quad (5)$$

XGBoost utiliza un enfoque de incremento progresivo para simplificar el modelo. En cada paso, cuando el modelo agrega un nuevo árbol de decisión, aprende una nueva función y sus coeficientes para ajustarse a los residuos predichos en la iteración anterior. De esta manera, en el paso  $t$ -ésimo, el valor predicho de  $x_i$  se calcula como  $Y_i^t = Y_i^{t-1} + f_t(x_i)$ . La función objetivo final se expresa en la ecuación 6.

$$Obj(\theta)^t = \sum_{i=1}^n l(y_i, \hat{y}_i^t) + \sum_{m=1}^t \Omega(f_m) = \sum_{i=1}^n l(y_i, \hat{y}_i^t + f_t(x_i)) + \sum_{m=1}^t \Omega(f_m) \quad (6)$$

Finalmente, XGBoost utiliza un enfoque "Greedy" al construir continuamente aprendices base para formar el modelo final.

**3.4.2. Bosque Aleatorio (Random Forest, RF).** Bosque aleatorio (Random Forest, RF, en inglés) es un algoritmo a través de la construcción de múltiples árboles de decisión para la clasificación de muestras de datos. Puede ser usado para problemas de regresión, clasificación y descomposición de componentes principales. Su idea principal es realizar estadística desde la división de datos en cada uno de sus árboles y seleccionar las categorías con mayor relación. Además, no se encuentra limitado por el tamaño de los datos, siendo capaz de soportar problemas que requieren dimensiones grandes de datos (Li et al., 2020). El algoritmo de RF genera inicialmente un número de árboles, y dentro de cada árbol genera unos nodos; luego de esto, se selecciona

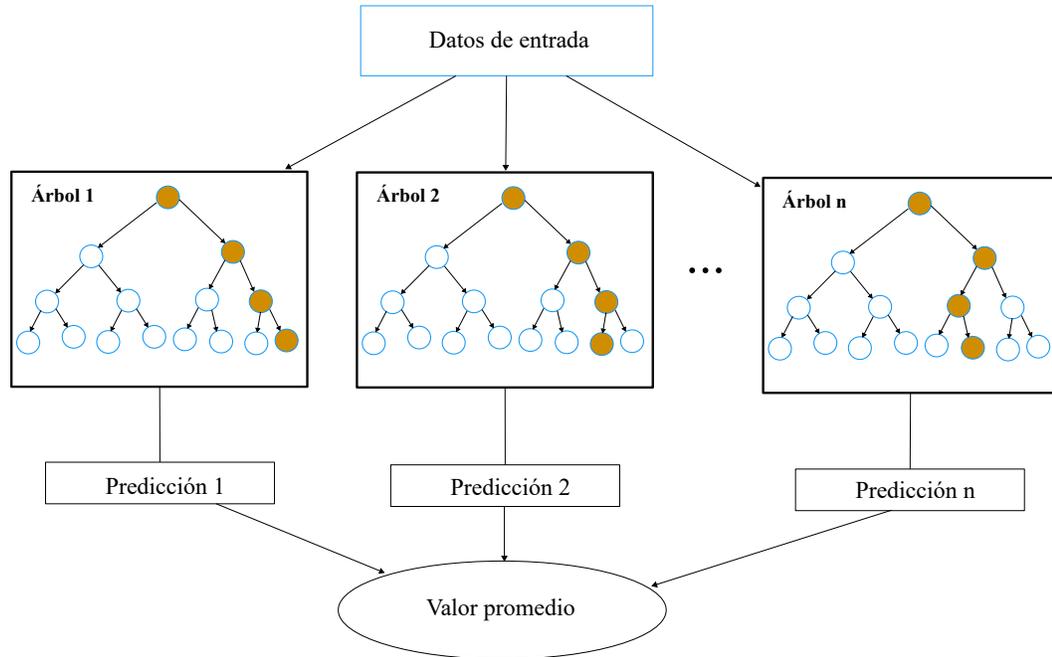
**Figura 3***Diagrama descriptivo del proceso del modelo XGBoost*

*Nota.* Figura ilustrativa de la metodología del algoritmo XGBoost. Adaptado de Anggraeni et al. (2021)

aleatoriamente una cantidad de características dentro de cada uno de los nodos y se calcula dentro de cada nodo un parámetro de ganancia. Al salir del nodo, se selecciona el nodo que obtuvo el mayor valor en ganancia para la salida del árbol y, finalmente se calcula el promedio de las salidas de cada uno de los árboles (Kumari et al., 2022), ver Figura 4.

RF es un algoritmo seleccionado debido a ser altamente recomendable para el tratamiento de problemas de predicción de demanda a corto plazo tal, e incluso puede obtener mejores desempeño que una red neuronal artificial (ANN) como evidencian los resultados de Prashanthi and Priyadarsini. (2021) y Wang et al. (2021), respectivamente.

**Figura 4**  
*Diagrama descriptivo del proceso del modelo Random Forest*



*Nota.* Figura ilustrativa de la metodología del algoritmo de Random Forest (RF). Adaptado de Kumari et al. (2022)

### 3.5. Métricas de evaluación

Principalmente, se escogieron dos métricas de precisión para evaluar el rendimiento de los modelos predictivos, el error porcentual absoluto medio (MAPE) y el error cuadrático medio (RMSE) definidos por las ecuaciones 7 y 8.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (8)$$

En las ecuaciones del MAPE y RMSE (7 y 8 respectivamente) N corresponde a la longitud de los datos de prueba. En cuanto a  $y_i$  y  $\hat{y}_i$ , corresponden a la demanda actual y predicha en cada paso de tiempo  $i$  respectivamente.

La elección de estas métricas de evaluación se basó en consideraciones clave dentro del campo de la predicción de demanda de energía eléctrica. El MAPE, ampliamente utilizado en el ámbito de modelos de aprendizaje automático, ofrece una medida intuitiva del porcentaje de error entre los valores predichos y los valores reales. Un MAPE más bajo indica una mayor proximidad entre las predicciones y los datos reales, lo que refleja un rendimiento mejorado en la capacidad predictiva del modelo a evaluar (Gao et al., 2020).

Por otro lado, el RMSE complementa esta perspectiva al proporcionar una medida cuantitativa en unidades de medición pertinentes, en este caso, megavatios (MW). Este enfoque cuantitativo captura la magnitud real de las desviaciones entre la demanda predicha y la demanda real, lo que resulta esencial para una comprensión más profunda de la precisión de las predicciones.

Esta combinación de métricas, el MAPE para evaluar la precisión porcentual y el RMSE para evaluar la magnitud de los errores, proporciona una evaluación completa y equilibrada de la calidad de las predicciones de demanda energética del modelo.

Finalmente, también fue contemplada la métrica de desviación horaria propuesta según el Artículo 30 de la Resolución CREG-060 de 2019 donde se clarifica que: "Se calculará el porcentaje

de desviación horaria de las plantas o unidades de generación como el valor absoluto de la diferencia de su despacho programado horario o el redespacho, según corresponda, y su generación real horaria, sobre su despacho programado horario o el redespacho, según corresponda."; lo que podría interpretarse matemáticamente como se muestra en la Ecuación 9. Esto se realiza con el fin de poder conocer el desempeño de los modelos en seccionamiento por tramos de veinticuatro horas y doce horas como se sugieren en la Resolución CREG-100 de 2019 y el Acuerdo CNO 1303 de 2020, correspondientemente.

$$Desviacion = \frac{|Pronostico - Real|}{Real} \cdot 100 \quad (9)$$

#### 4. Análisis de resultados

Para la evaluación de rendimiento serán calculadas las métricas de error MAPE y RMSE resultantes del pronóstico de los modelos de aprendizaje de máquina propuestos, los cuales utilizan de entrada los escenarios de predictores resultantes de la aplicación de los algoritmos RFE, SFS y SBS según lo registrado en la Tabla 3, Tabla 4 y Tabla 5, respectivamente.

La evaluación se lleva a cabo para dos escenarios de predicción. En el primer escenario, llamado Caso 1, se emplean datos reales para el horizonte de pronóstico de pruebas (julio 2023 - agosto 2023). Se parte del supuesto de que al final de cada día se cuentan con los datos de consumo correspondientes a ese día, y se realiza un reentrenamiento semanal del algoritmo. En el segundo escenario, denominado Caso 2, se emplean los datos pronosticados por el modelo para estimar la demanda dentro del horizonte de pronóstico de pruebas. Este enfoque se conoce en la literatura como pronóstico continuo (Rolling forecasting, en inglés). Al igual que en el primer caso, se realiza un reentrenamiento semanal del algoritmo con datos reales, asumiendo que al final de cada semana se disponen de los datos de demanda energética.

En función de los resultados, se seleccionarán los escenarios más destacados a los cuales se les aplicarán las evaluaciones requeridas por la legislación colombiana, según las regulaciones CNO 1303 de 2020 y CREG 100 de 2019. Además, se explorarán las limitaciones del enfoque de pronóstico continuo utilizado en el Caso 2 para los mejores escenarios, proponiendo en su lugar una metodología alternativa.

#### 4.1. Predicción con datos reales (Caso 1)

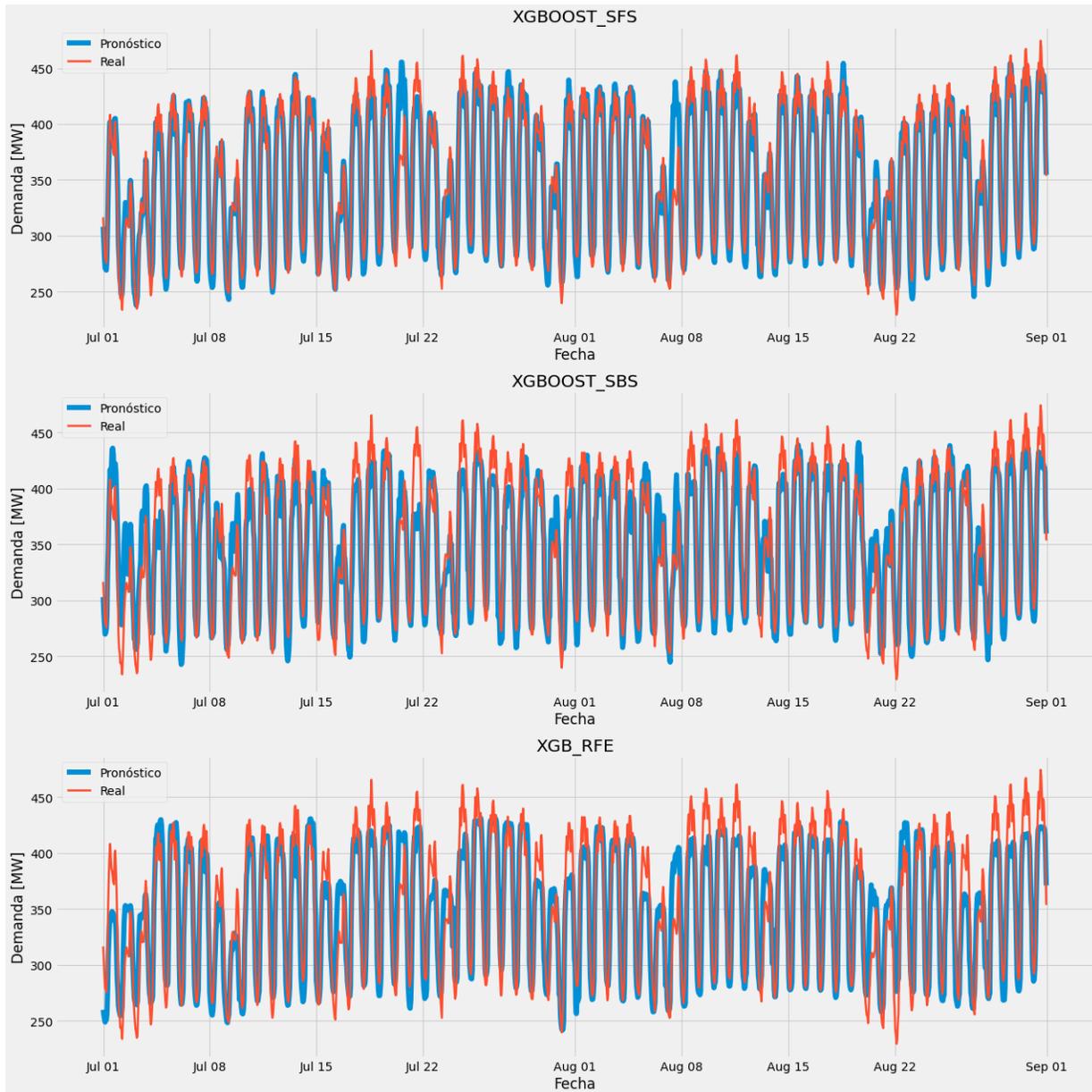
En el primer enfoque, se partió del supuesto de contar con datos de demanda energética reales al final de cada día. Teniendo en cuenta que el modelo pronostica la demanda del día  $n$ , utilizando información extraída de la demanda los seis días previos al día objetivo ( $n - 1, n - 2, \dots, n - 6$ ). Dado que el operador de red (OR) realiza el pronóstico semanal, este enfoque es una situación ideal en el contexto de predicción de demanda energética en el departamento. No obstante, la relevancia de este enfoque radica en la capacidad de proporcionar una visión precisa de la habilidad y el rendimiento del algoritmo en la predicción a un horizonte temporal muy corto, es decir, apenas 24 horas. Esto permitió evaluar su desviación con la demanda energética real y las predicciones realizadas por el OR.

Una comparativa entre el pronóstico obtenido en el Caso 1 y la demanda real, se muestra en las Figuras 5 y 6, con los modelos basados en XGBoost y Random Forest respectivamente.

Los errores obtenidos por las métricas de evaluación MAPE y RMSE se presentan en la Tabla 6. En las representaciones de los resultados es evidente que el comportamiento general de las diversas combinaciones de métodos de aprendizaje de máquina y selección de predictores tiene un rendimiento competente, sin embargo, es destacable el rendimiento superior obtenido por las combinaciones XGB-SFS y RF-SFS. Así, para un horizonte de predicción de 24 horas, los modelos destacados tienen un rendimiento bastante satisfactorio, superando las predicciones realizadas por el operador de red, con una desviación media menor de  $1,1512[MW]$  para XGB-SFS y  $0,4898[MW]$  para RF-SFS. En la Figura 7 se observa el comportamiento del MAPE promedio para cada día de

**Figura 5**

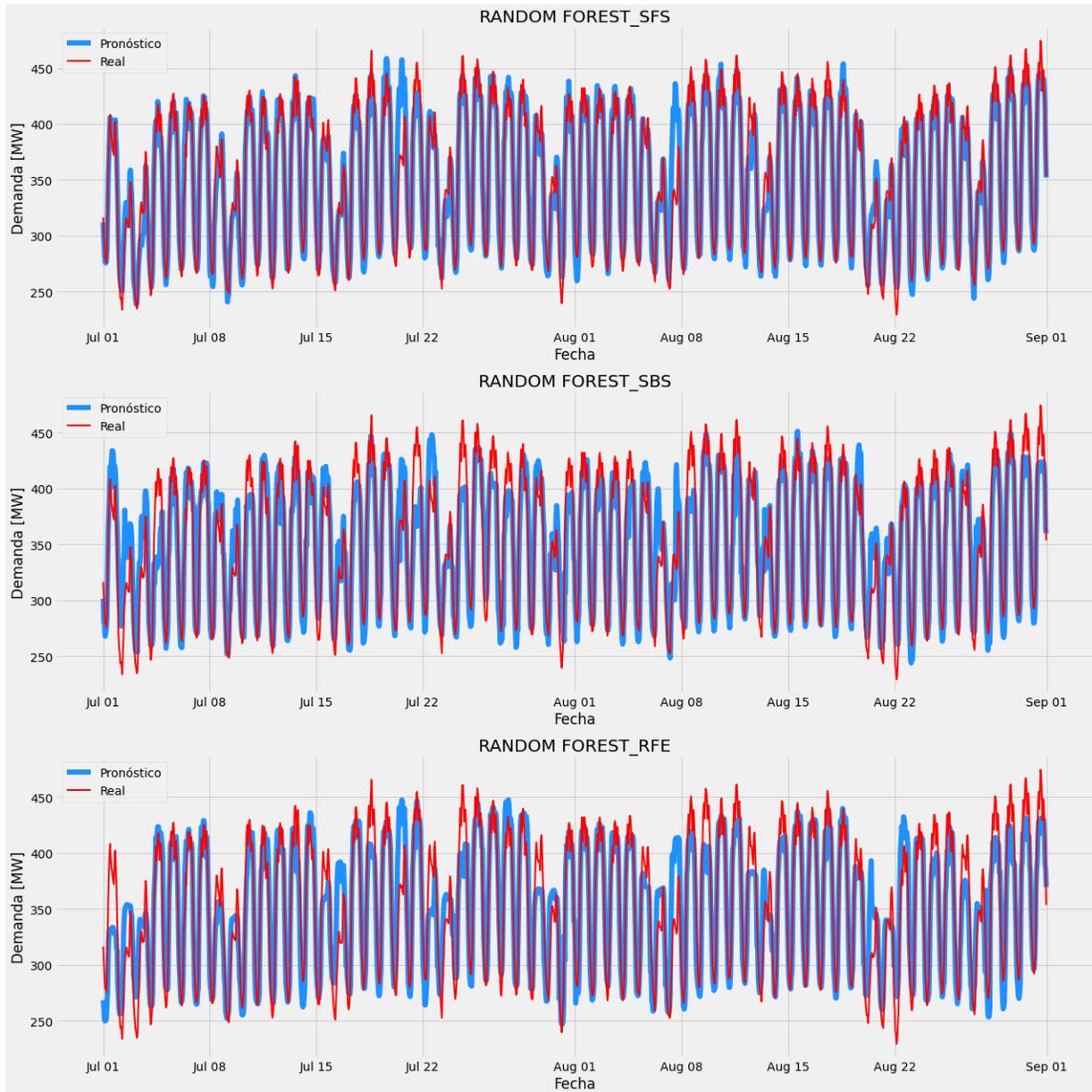
Gráfica en el tiempo comparativa entre lo pronosticado por el modelo XGBoost y la demanda real para los meses julio y agosto del año 2023 (Caso 1).



la semana. Del mapa de calor se puede evidenciar que en general las combinaciones tienen un MAPE mayor los días domingo y lunes, siendo los únicos casos donde las dos combinaciones

**Figura 6**

Gráfica en el tiempo comparativa entre lo pronosticado por el modelo RF y la demanda real para los meses julio y agosto del año 2023 (Caso 1).



destacadas no tienen un rendimiento superior al del OR. Para el resto de días, el comportamiento es aceptable, salvo en el caso de las combinaciones que utilizan el método RFE de selección de

predictores, donde el error se acentúa también los fines de semana, algo inesperado, esto muy probablemente porque la máxima cantidad de días hacia atrás que observan los modelos (según los predictores) bajo una función de modo intrínseco (IMF) o demanda neta es de cuatro (4) días hacia atrás. Por esta razón, aunque se contemplen los festivos como predictor quizás no existe para el modelo de aprendizaje una referencia muy clara de tendencia semanal de consumo energético al llegar el fin de semana; un análisis análogo que podría apoyar la hipótesis propuesta se da al revisar los predictores resultantes del método de selección SBS que, aunque no se contemple festivos, sí alcanza a conocer mejor la tendencia semanal de consumo debido a que su panorama de visión es hasta de seis (6) días antes, lo que conlleva a que el fin de semana (sobre todo el día sábado) haya mejores resultados bajo la métrica MAPE en comparación a RFE.

Pese al buen rendimiento obtenido por ambos métodos, este sigue siendo un caso netamente académico, y distante de la forma práctica de pronosticar demanda energética realizada por el OR, de horizonte semanal, razón por la cual se propone el Caso 2.

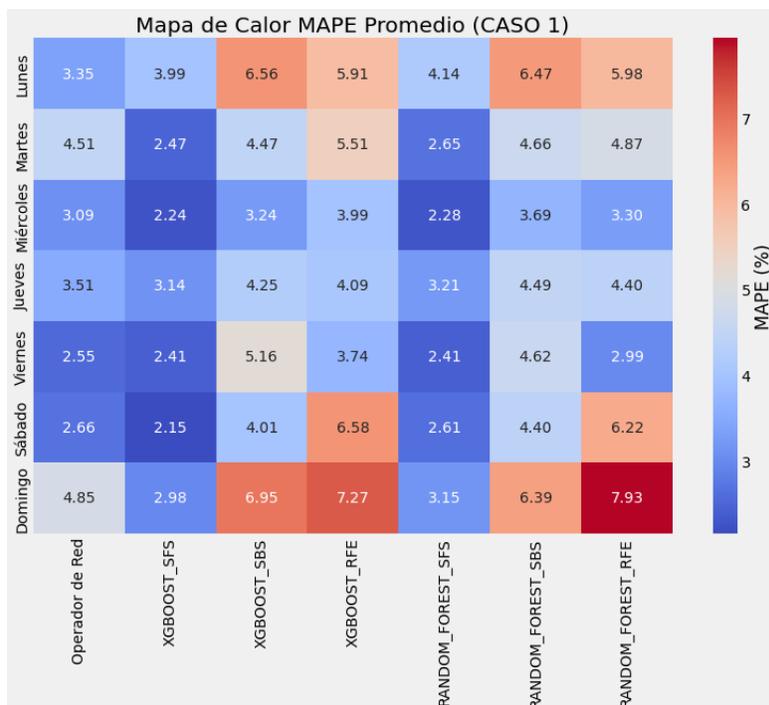
#### **4.2. Predicción con datos pronosticados (Caso 2)**

El segundo caso propuesto busca ser lo más similar posible a cómo realmente el operador de red (OR) hace entrega del pronóstico final de la demanda al Centro Nacional de Despacho (CND); el reporte de pronóstico mencionado contiene el posible consumo energético que habrá hora a hora desde el día lunes hasta el día domingo de esa semana.

Para recrear este escenario en la predicción, se parte del histórico de datos para predecir el primer día de la semana, se almacena el resultado de esta predicción y se realiza el pronóstico del segundo día, usándose una metodología de predicción en cascada hasta el último día de la semana;

**Figura 7**

Mapa de calor, métrica MAPE diaria para comparativa entre el Operador de Red (OR) y los modelos propuestos en la predicción con datos reales (Caso 1)



para finalizar, luego de haber sido realizada la predicción de la semana, se reentrena el algoritmo asumiendo la existencia de los datos de consumo reales a final de semana. Este enfoque recursivo de predicción se conoce como 'pronóstico continuo', donde se emplea un modelo de predicción de un solo paso en múltiples iteraciones. En este caso, la predicción del paso actual, denotada como  $n_t$ , se utiliza como entrada para prever el siguiente paso,  $n_{t+1}$ . Esta metodología de pronóstico es reconocida en la literatura especializada en predicción de series de tiempo Li (2020), Deng et al. (2022). Así mismo, este enfoque fue previamente utilizado en el contexto de predicción de demanda energética en el departamento de Santander por Agudelo and Barón (2022).

A pesar de sus ventajas, el pronóstico continuo, al utilizar predicciones como datos de

**Tabla 6***Recopilación de desempeño para los modelos en Caso 1*

<b>Métrica MAPE y RMSE para los distintos modelos evaluados (Caso 1)</b>		
<b>Modelo</b>	<b>MAPE (%)</b>	<b>RMSE</b>
OR	3,5172	15,1307
RF_SBS	4,9661	23,2533
RF_RFE	5,1334	23,9662
RF_SFS (*)	2,9299	14,6409
XGB_SBS	4,9465	22,2778
XGB_RFE	5,3326	23,7257
XGB_SFS (*)	2,7738	13,9795

*Nota.* Los modelos que cuentan con un (\*) son aquellos dos que contaron con un mejor desempeño en según las métricas MAPE y RMSE para Caso 1.

entrada, puede dar lugar a la acumulación de errores de predicción, lo que puede afectar la precisión del modelo a medida que se aumenta el horizonte de pronóstico. Por esta razón, en la siguiente subsección, se analizan las limitaciones de este enfoque y se propone una metodología adicional para los mejores escenarios obtenidos en el Caso 2.

Con este caso se busca evaluar cómo se comportarían los modelos propuestos en un escenario realista, y, así mismo, realizar una comparativa de rendimiento con respecto a los resultados que ha obtenido el operador de red en el mismo lapso de tiempo bajo esta modalidad.

En las Figuras 8 y 9, se presenta una comparativa entre las predicciones realizadas para el Caso 2 y la demanda real utilizando modelos basados en XGB y RF, respectivamente. En comparación al caso ideal, es evidente la diferencia de rendimiento entre las combinaciones de modelos, destacando el desempeño superior de aquellos modelos entrenados con los predictores seleccionados mediante SFS (Sequential Feature Selection). Para una evaluación cuantitativa de

**Tabla 7***Recopilación de desempeño para los modelos en Caso 2*

<b>Métrica MAPE y RMSE para los distintos modelos evaluados (Caso 2)</b>		
<b>Modelo</b>	<b>MAPE (%)</b>	<b>RMSE</b>
OR	3,5172	15,1307
RF_SBS	9,7763	46,3365
RF_RFE	5,8752	27,0106
RF_SFS (*)	3,1165	15,2585
XGB_SBS	10,2435	47,8059
XGB_RFE	6,38	28,6242
XGB_SFS (*)	2,8976	14,5884

*Nota.* Los modelos que cuentan con un (\*) son aquellos dos que contaron con un mejor desempeño en según las métricas MAPE y RMSE para Caso 2.

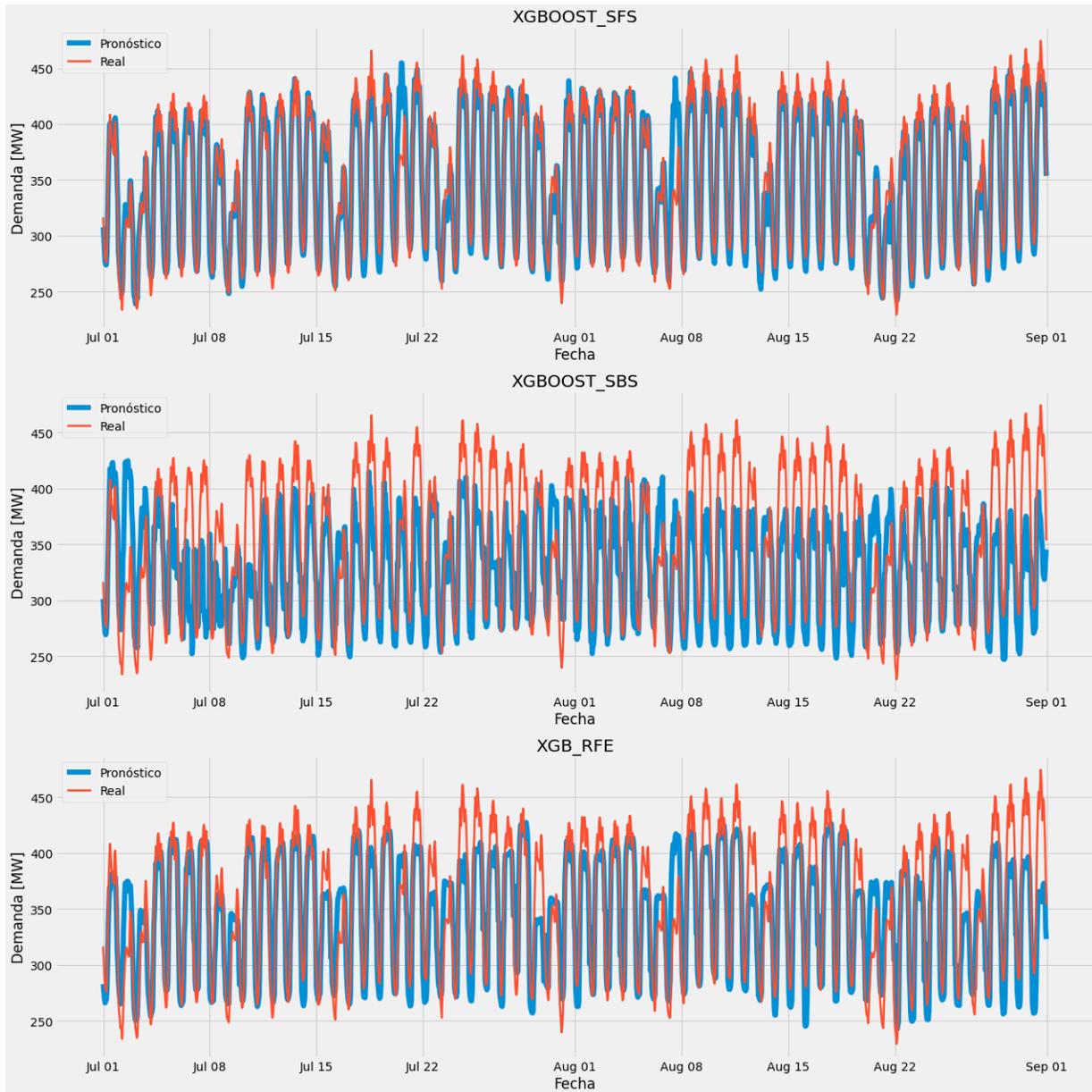
estas diferencias de rendimiento, se presentan los resultados de las métricas de evaluación en la Tabla 7. Es notable el mejor rendimiento en el MAPE de los modelos entrenados con los predictores del grupo SFS. En ambos casos, estos modelos superan el desempeño del modelo utilizado por el operador de la red, destacando el modelo XGB-SFS como el mejor en ambas métricas de error.

En la Figura 10, se presenta un mapa de calor que muestra el MAPE promedio para cada día de la semana. Al igual que en las representaciones previas de los resultados, se hace evidente el desempeño superior de los modelos entrenados con SFS, ya que superan a los demás modelos, incluyendo al utilizado por el OR (Operador de Red) en todos los días de la semana, con la excepción del lunes, donde el MAPE registra un incremento significativo en comparación con los demás días para los modelos XGB-SFS y RF-SFS.

Una posible mejora para el modelo sería considerar el lunes como un tipo de día especial y tratarlo de manera distinta en la selección de predictores, de manera similar a cómo se abordan

**Figura 8**

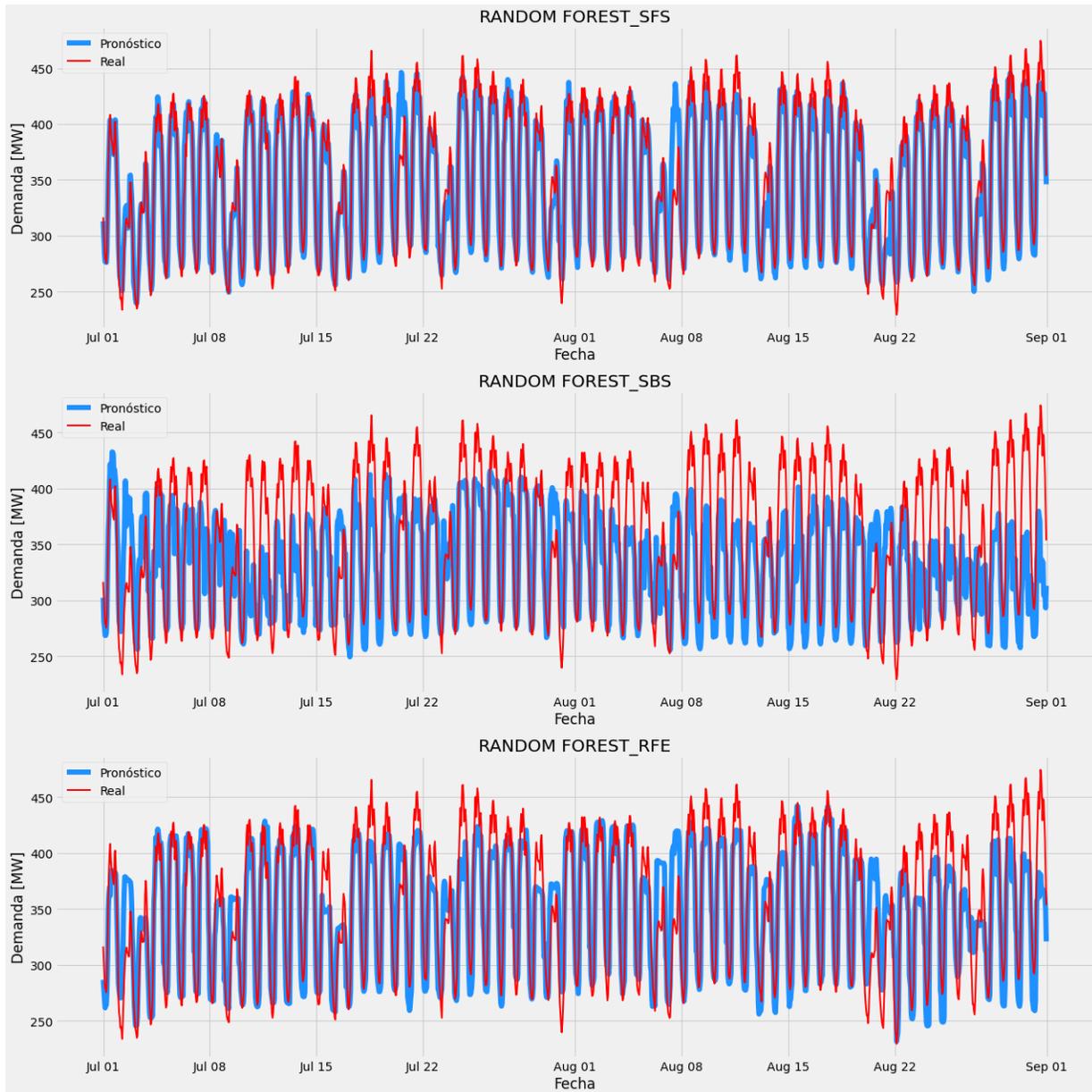
Gráfica en el tiempo comparativa entre lo pronosticado por el modelo XGBoost y la demanda real para los meses julio y agosto del año 2023 (Caso 2).



los días festivos en Colombia y los fines de semana. Esto podría ayudar a mejorar aún más el rendimiento del modelo en estas circunstancias particulares.

**Figura 9**

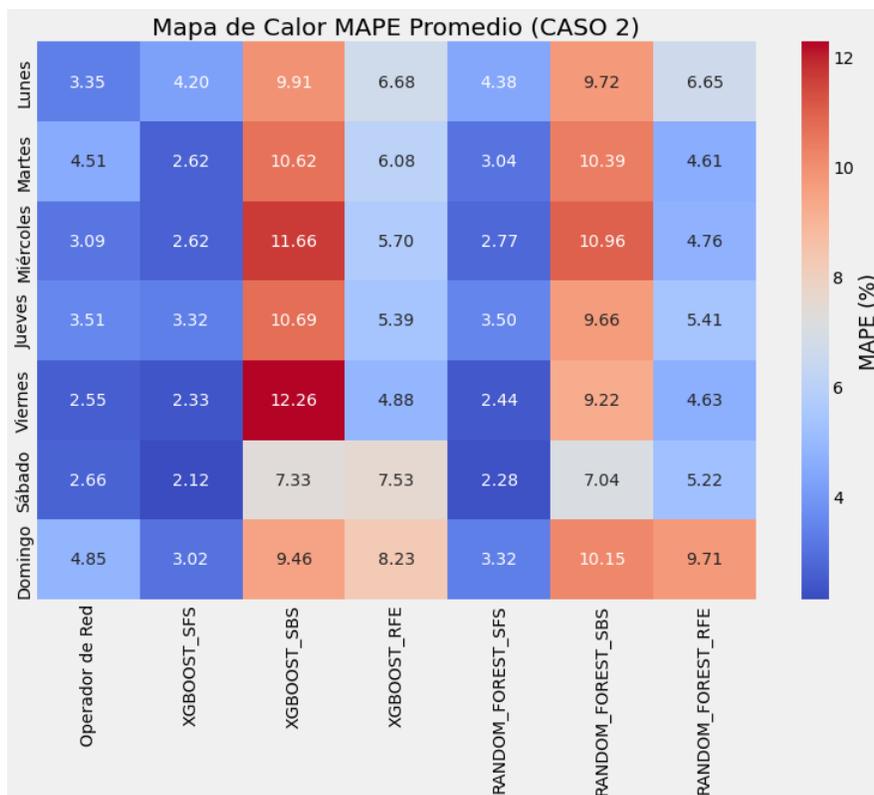
Gráfica en el tiempo comparativa entre lo pronosticado por el modelo RF y la demanda real para los meses julio y agosto del año 2023 (Caso 2).



Como se mencionó anteriormente, los modelos que utilizan Sequential Feature Selection (SFS) exhiben un rendimiento significativamente superior en comparación con las combinaciones

**Figura 10**

Mapa de calor, métrica MAPE diaria para comparativa entre el Operador de Red (OR) y los modelos propuestos en la predicción con datos pronosticados (Caso 2)



que emplean Recursive Feature Elimination (RFE) y Sequential Backward Selection (SBS). Al examinar los predictores seleccionados por cada método, los cuales se presentan en las Tablas 3, 4 y 5, se observó que tanto RFE como SBS principalmente eligen predictores relacionados con la demanda energética de los días previos al día que se está prediciendo, con la única excepción de la inclusión de los festivos por parte de RFE.

Es importante destacar que, al realizar una comparación directa entre RFE y SBS, cuyas predicciones se basan prácticamente en datos relacionados con la demanda, RFE exhibe un rendimiento notablemente superior, ya sea utilizando XGB o RF. A diferencia de SBS, RFE incorpora

un dato calendario (festivos). Esto subraya la importancia crítica de los datos de tipo calendario en la predicción de la demanda energética en el departamento de Santander.

Por otro lado, los predictores seleccionados por SFS toman en cuenta el patrón cíclico del consumo de energía al incluir la hora, los días festivos y el día de la semana. Esta elección, combinada con el rendimiento superior de RFE en comparación con SBS al considerar solo los festivos, subraya aún más la relevancia de los datos del calendario en las predicciones de la demanda energética.

Además, es destacable que SFS también incorpora la variable meteorológica temperatura en sus selecciones. La inclusión temperatura es indicador de su importancia en las predicciones de demanda energética y resalta la capacidad de SFS para considerar múltiples aspectos relevantes en su proceso de selección de características.

#### **4.3. Mejores escenarios: Comparativa respecto al Operador de Red (OR) y cálculo de desviaciones**

En esta sección se llevará a cabo una revisión detallada de las limitaciones inherentes al enfoque de pronóstico continuo, que fue empleado en el Caso 2, mediante el análisis estadístico de los resultados obtenidos por los modelos destacados. Se utiliza el Caso 2 debido a que se trata de una metodología similar a como lo realiza el Operador de Red (OR). La selección de los mejores modelos toma en cuenta los resultados del Caso 1 y Caso 2, por lo tanto, para este caso los modelos analizados son XGB-SFS y RF-SFS. Para abordar estas limitaciones en futuros trabajos en esta área, se propone una metodología alternativa de pronóstico que, en conjunto con un modelo de múltiples etapas, podría solventar de manera efectiva las deficiencias del enfoque de pronóstico

continuo. Además, se realiza una comparativa de desempeño entre los dos métodos más destacados que surgieron del análisis previo, en conformidad con las regulaciones principales establecidas por la Comisión de Regulación de Gas y Energía (CREG) en su Resolución 100 de 2019 y el Acuerdo 1303 de 2020 propuesto por el Centro Nacional de Operación del Sector Eléctrico (CNO). Esta evaluación se centra en la desviación diaria en tramos de veinticuatro (24) horas y de doce (12) horas, respectivamente.

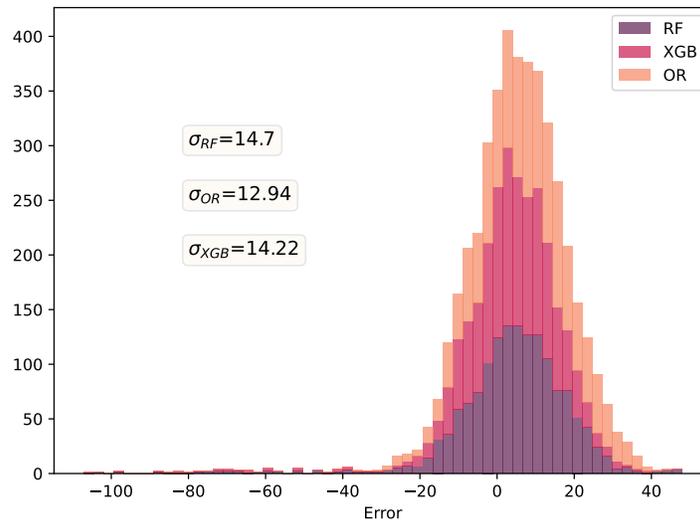
**4.3.1. Análisis estadístico para modelos destacados.** Como se mencionó en la sección correspondiente al Caso 2, al utilizar predicciones como datos de entrada del modelo, el pronóstico continuo da lugar a la acumulación de errores de predicción a lo largo del tiempo, por lo que a medida que aumente el horizonte de pronóstico, la precisión del modelo pueden disminuir. Por esta razón se llevó a cabo un análisis estadístico detallado del comportamiento del error de los dos mejores casos (XGB-SFS y RF-SFS) para el Caso 2, y se comparó con los resultados obtenidos por el Operador de Red (OR). En la Figura 11 se muestra la distribución normal del error obtenido por los dos casos destacados y el OR.

De la Figura 11, se puede observar que, a pesar de que los resultados absolutos obtenidos por los mejores modelos superan en rendimiento al del Operador de Red (OR) (Tabla 7), la desviación estándar del error para los modelos destacados es mayor. Esto sugiere que los modelos destacados son menos precisos que el que utiliza el OR. Esta discrepancia se debe a la acumulación gradual del error a lo largo del tiempo debido a la técnica de pronóstico continuo, que aumenta a medida que transcurre la semana de pronóstico.

Para analizar por qué, a pesar de esto, los modelos destacados presentan un mejor rendi-

**Figura 11**

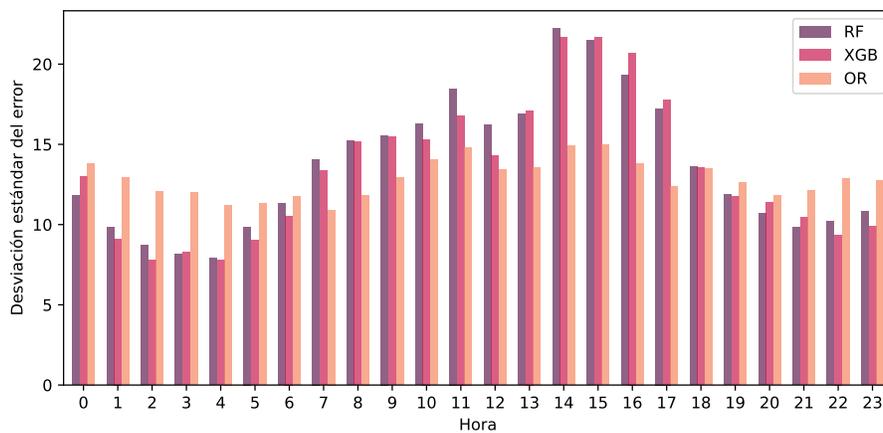
*Distribución normal del error para los modelos destacados y operador de red (OR)*



miento bajo las métricas MAPE y RMSE, se examinó el comportamiento horario de la desviación estándar del error, como se muestra en la Figura 12.

**Figura 12**

*Histograma de desviación estándar hora-hora para los modelos destacados y operador de red (OR)*



De la Figura 12 se puede evidenciar que los modelos destacados tienen un rendimiento

similar, obteniendo una desviación estándar del error menor a la obtenida por el OR durante las horas 0 a 7 y de 18 a 23. Esto es particularmente relevante dado que las horas de mayor demanda de potencia eléctrica abarcan desde las 18 hasta las 23 horas como se evidencia en Agudelo and Barón (2022) (Esta información acerca de las horas de mayor consumo también puede ser consultada en el portal web de XM S.A ESP). Este hecho explica por qué, a pesar de que los modelos destacados presentan una desviación estándar más alta debido al error acumulado por el pronóstico continuo, superan al OR en términos de las métricas MAPE y RMSE, lo que se traduce en una predicción global de la demanda de mayor calidad.

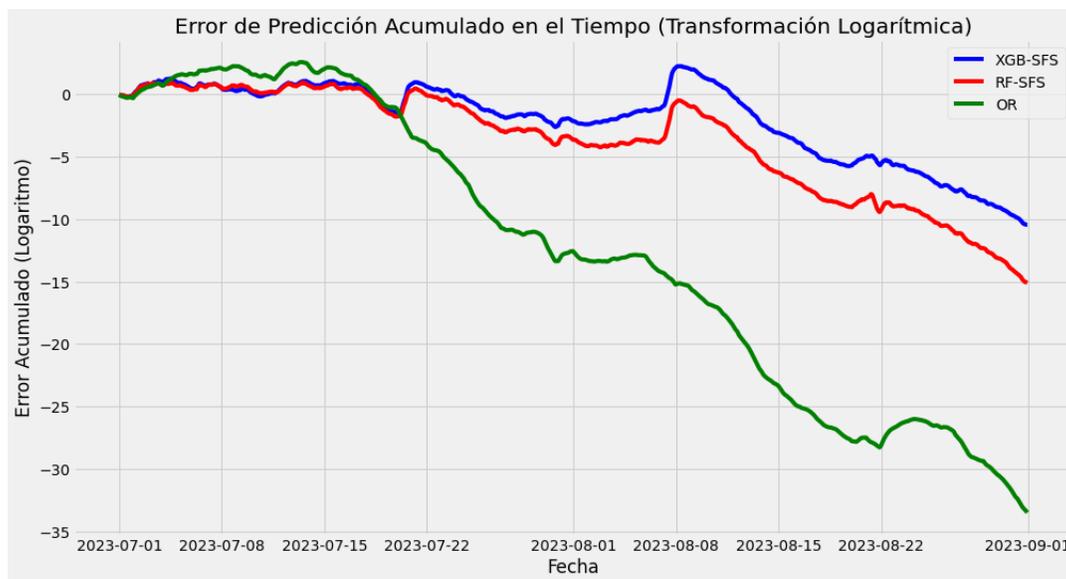
En la Figura 13 se puede constatar la predicción global más acertada de los modelos destacados respecto al OR, mostrándose el comportamiento del error acumulado en el tiempo, en una escala logarítmica, para los dos mejores modelos y el operador de red (OR). En la Ecuación 10 se evidencia la expresión matemática utilizada con cada uno de los resultados de los modelos que posteriormente fueron graficados.

$$Error_{acumulado} = \sum \log\left(\frac{Pronostico}{Real}\right) \quad (10)$$

También es importante destacar la relación presente en la Ecuación 10. Un error acumulado mayor, al ser el logaritmo de un cociente, equivale a la resta del logaritmo de la predicción y el logaritmo de la demanda real. Un error acumulado más negativo indica que las predicciones están por debajo del valor de la demanda real, mientras que un error cercano a cero o incluso positivo indica no solo predicciones más precisas, sino también una sobre-estimación de la demanda por

**Figura 13**

*Cálculo de error acumulado con los mejores modelos y el operador de red en el tiempo (en escala logarítmica)*



parte del algoritmo.

**4.3.2. Metodología alternativa de ventana ampliada.** Dado el reconocimiento de las limitaciones inherentes al enfoque empleado en el pronóstico del Caso 2, se propone una metodología que replicará el proceso real utilizado por el Operador de Red (OR) para proporcionar el pronóstico final al Centro Nacional de Despacho (CND). Esta propuesta tiene como objetivo abordar de manera efectiva las debilidades del enfoque de pronóstico continuo. En trabajos futuros en este campo, se espera que la implementación de esta metodología de pronóstico, en conjunto al enfoque de múltiples etapas desarrollado en este estudio, resulte en una mejora sustancial en la precisión general de la predicción de la demanda energética para el departamento de Santander.

En este enfoque de pronóstico, en lugar de utilizar la demanda de los 6 días previos al día a predecir como entradas para el espacio inicial de información, se optó por utilizar la demanda de

una a cuatro semanas anteriores al día objetivo de pronóstico. Esta decisión se basa en la suposición de que, al final de cada semana, el Operador de Red (OR) dispone de datos reales sobre el consumo energético del departamento. De esta manera, se puede predecir una ventana de tiempo de una semana utilizando información de la demanda energética del mes anterior a esa semana específica. Esto permite evitar utilizar valores pronosticados como entradas en el algoritmo, lo que a su vez ayuda a prevenir la acumulación de errores de pronóstico a lo largo del tiempo.

Teniendo el espacio de información inicial definido, el resto del proceso se realiza similar al propuesto, a excepción de que se utiliza únicamente un método de selección de predictores, SFS en este caso, escogido por su desempeño en los resultados obtenidos previamente. Los predictores escogidos por SFS con el nuevo espacio de información se evidencian en la Tabla 8. Se aprecia que, al igual que con el espacio de información anterior, son tenidas en cuenta las variables calendario, y se resalta que otra variable meteorológica, la presión, entra a formar parte de las entradas de la máquina.

Una vez definido el espacio de información inicial, el resto del proceso se lleva a cabo de manera similar a la propuesta anterior, con la única diferencia de que en este caso se utilizó exclusivamente el método de selección de predictores Selección Secuencial hacia Adelante (SFS), elegido debido a su buen desempeño en los resultados previamente obtenidos. La Tabla 8 muestra los predictores seleccionados por SFS con el nuevo espacio de información. Es importante destacar que, al igual que en el espacio de información anterior, se incluyen las variables de calendario. Además, es relevante señalar que otra variable meteorológica, la presión atmosférica, se ha incorporado como parte de las entradas para el modelo.

**Tabla 8**

*Predictores seleccionados por el algoritmo de selección secuencial de características hacia al frente (SFS) para un horizonte de una semana (Ventana ampliada, Caso 2 - VA)*

<b>Predictores escogidos por el algoritmo SFS para el horizonte de ventana ampliada (Caso 2 - VA)</b>	
<b>Predictor</b>	<b>Acrónimo</b>
Hora (En formato categórico, de la hora 0 a la 23 del día)	hour
Festivos	holiday
Temperatura (en grados Celsius)	Temp
Presión (en bar)	Pressure
Día (En formato categórico, del día 0 al día 6)	t5
Demanda una semana antes del día evaluado	demand_1_weeks_before
Demanda tres semanas antes del día evaluado	demand_3_weeks_before
Segunda función de modo intrínseco para la demanda cuatro semanas antes	demand_4_weeks_before_IMF_2
Tercera función de modo intrínseco para la demanda cuatro semanas antes	demand_4_weeks_before_IMF_3
Cuarta función de modo intrínseco para la demanda cuatro semanas antes	demand_4_weeks_before_IMF_4

La Tabla 9 presenta una comparación de los resultados obtenidos en términos de las métricas MAPE y RMSE entre los dos mejores modelos del Caso 2, el Operador de Red (OR), y el enfoque propuesto en esta sección que, por facilidad, se denominará como Caso 2 - VA (Ventana Ampliada).

Aunque se observa una ligera disminución en el rendimiento del Caso 2 - VA en comparación con el Caso 2, el modelo exhibe un buen desempeño en general. Para optimizar aún más su rendimiento, es necesario adaptar el método de predicción propuesto a las características específicas del departamento de Santander. Este enfoque representa una perspectiva interesante para investigaciones futuras.

En la Figura 14, se representa la distribución normal del error para los modelos de pronóstico en el Caso 2 - VA, así como para el Operador de Red (OR). A pesar de evitar problemática relacionada con el enfoque de pronóstico continuo presente en el Caso 2, se observa que la desvia-

**Tabla 9**

*Recopilación de desempeño para los mejores modelos para Caso 2, Caso 2 - VA y OR*

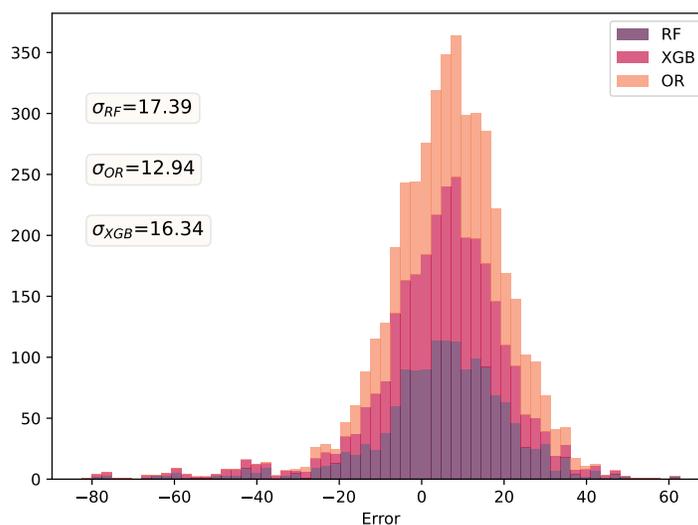
<b>Métrica MAPE y RMSE para los distintos modelos evaluados (Comparativo C2, C2-VA y OR)</b>		
<b>Modelo</b>	<b>MAPE (%)</b>	<b>RMSE</b>
OR	3,5172	15,1307
RF_SFS_C2	3,1165	15,2585
XGB_SFS_C2	2,8976	14,5884
RF_SFS_C2VA	3,7610	18,0402
XGB_SFS_C2VA	3,5674	16,8874

*Nota.* Los seudónimos C2 y C2VA que se encuentran en conjunto a los nombres de los modelos hacen referencia a Caso 2 y a la metodología de ventana ampliada (C2 - VA), respectivamente.

ción estándar del error empeoró para ambos modelos del Caso 2 - VA. No obstante, es fundamental tener en cuenta que no se llevó a cabo un estudio exhaustivo en la selección del espacio de información inicial para este enfoque de predicción, adaptándolo al contexto de Santander.

**Figura 14**

*Histograma de desviación estándar hora-hora para los modelos destacados y operador de red (OR) resultante de metodología de ventana ampliada (Caso 2 - VA)*

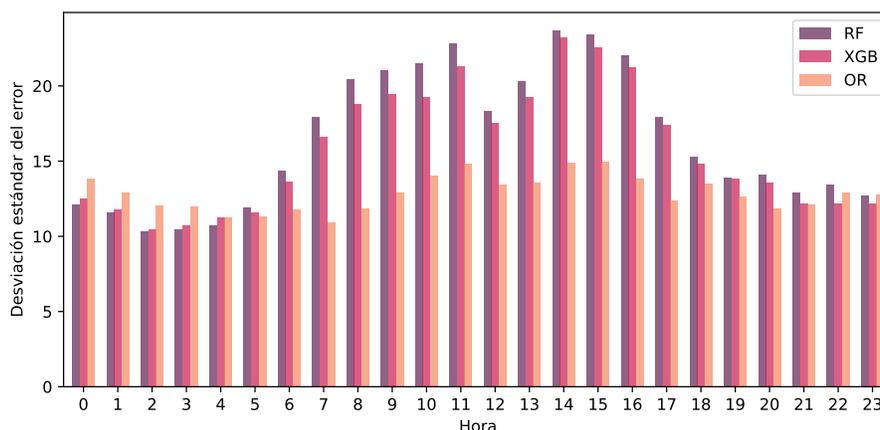


Para profundizar en el análisis de la disminución de rendimiento, se elaboró un histograma

que muestra la desviación estándar del error por horas, presentado en la Figura 15.

### Figura 15

*Distribución normal del error para los modelos destacados y operador de red (OR) resultante de metodología de ventana ampliada (Caso 2 - VA)*



Es evidente que, a diferencia del Caso 2, el Caso 2 - VA muestra un rendimiento inferior durante las horas pico de consumo energético del departamento, que son cruciales para evaluar el desempeño de los modelos del Caso 2. Esto se debe a que el Caso 2 considera días previos al día que se está pronosticando, lo que le permite capturar patrones de consumo energético de la semana en curso. En contraste, el Caso 2 - VA se basa en semanas anteriores a la semana que se está pronosticando. Este aspecto debe ser considerado en futuras investigaciones al seleccionar las posibles entradas para el modelo, siguiendo un enfoque similar al propuesto en el Caso 2 - VA.

**4.3.3. Desviaciones según Acuerdo CNO 1303 de 2020.** El cálculo de la desviación para evaluar las métricas indicadas en las regulaciones fue realizado exclusivamente con los resultados obtenidos al ser aplicado el Caso 2, debido a que fue el modelo con mejor desempeño entre las metodologías propuestas y ser similar a como realiza el pronóstico el Operador de Red

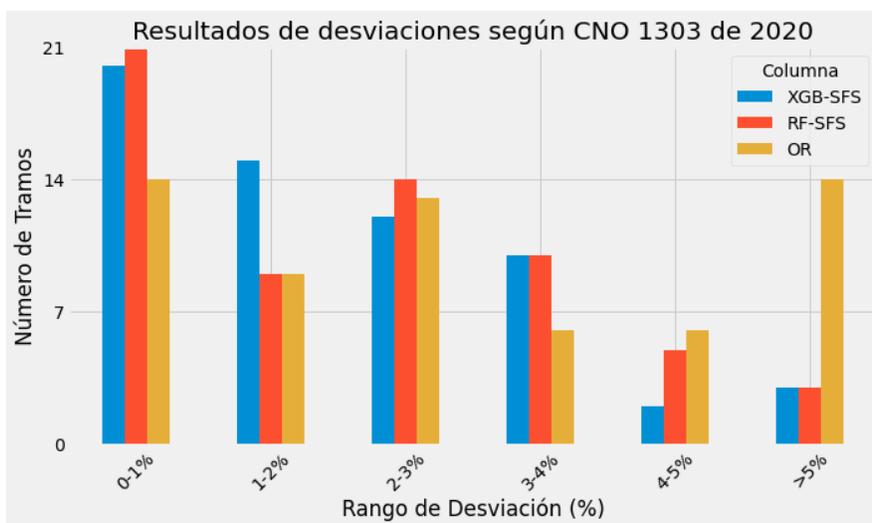
(OR).

En concordancia con las directrices del Acuerdo CNO 1303 de 2020, se procede al cálculo de las desviaciones para un total de 62 tramos. Estos tramos se corresponden con la demanda neta diaria de los meses de agosto y julio. Los resultados detallados, que incluyen el número de tramos dentro de cada rango de porcentaje de desviación para los dos modelos más destacados y el operador de red, se encuentran resumidos en la Tabla 10. Es relevante destacar que en comparación con la evaluación realizada por el Operador de Red (OR), se logró una reducción considerable del 79% en el número de tramos que podrían estar sujetos a penalizaciones según lo establecido en el mencionado acuerdo. Este resultado resalta la eficacia de ambos modelos en la predicción de desviaciones. La Figura 16 da una representación gráfica de los resultados mencionados.

El listado de resultados de desviaciones con los que fueron realizadas las tablas y gráficas anteriormente mencionadas es posible visualizarlas en el Apéndice B de este documento.

### Figura 16

*Gráfica de barras para el número de tramos por rangos de desviación según resultados bajo el Acuerdo CNO 1303 de 2020.*



**Tabla 10**

*Tabla recopilatoria del número de tramos por rangos de desviación según resultados bajo el Acuerdo CNO 1303 de 2020.*

Número de días en cada rango			
Rango	XGB-SFS	RF-SFS	OR
0 - 1 %	20	21	14
1 - 2 %	15	9	9
2 - 3 %	12	14	13
3 - 4 %	10	10	6
4 - 5 %	2	5	6
>5 %	3	3	14

**4.3.4. Desviaciones según Resolución CREG 100 de 2019.** En el contexto de la Resolución CREG 100 de 2019, se procedió al cálculo de las desviaciones para un total de 124 tramos. Estos tramos corresponden a la demanda de las primeras 12 horas del día y las siguientes 12 horas del día, para todos los días de la semana durante los meses de agosto y julio. Los resultados detallados, que incluyen el número de tramos dentro de cada rango de porcentaje de desviación para los dos modelos más destacados y el operador de red, se presentan en la Tabla 11. Es relevante destacar que en comparación con la evaluación realizada por el operador de red (OR), se observa una reducción significativa del 65 % y 56 % en el número de tramos que podrían estar sujetos a penalizaciones según lo establecido en la mencionada resolución. Este resultado resalta la eficacia de los modelos XGB-SFS y RF-SFS en la predicción de desviaciones. La Figura 17 da una representación gráfica de los resultados mencionados.

De la misma manera, es posible observar las desviaciones calculadas día a día en el Apéndice C del presente documento.

**Figura 17**

Gráfica de barras para el número de tramos por rangos de desviación según resultados bajo la Resolución CREG 100 de 2019.

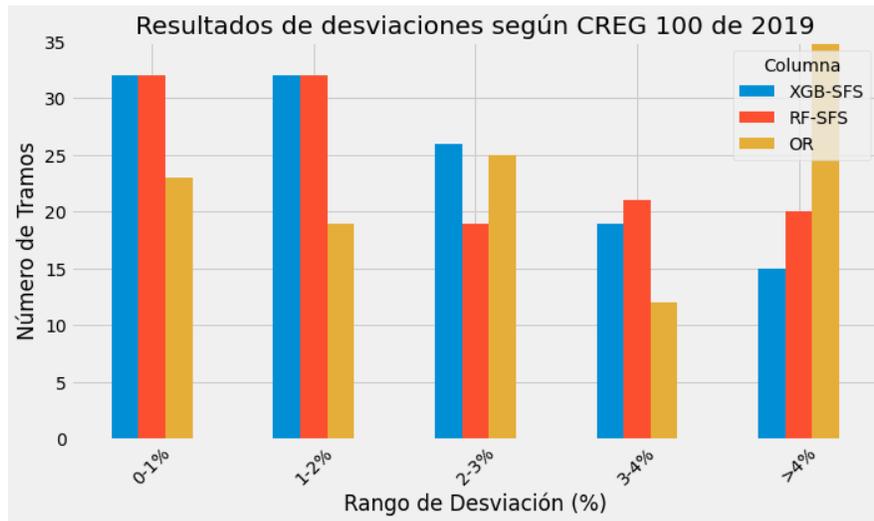
**Tabla 11**

Tabla recopilatoria del número de tramos por rangos de desviación según resultados bajo la Resolución CREG 100 de 2019.

Número de días en cada rango			
Rango	XGB-SFS	RF-SFS	OR
0 - 1 %	32	32	23
1 - 2 %	32	32	19
2 - 3 %	26	19	25
3 - 4 %	19	21	12
>4 %	15	20	45

## 5. Conclusiones y recomendaciones a futuro

A lo largo de este trabajo se ha destacado la importancia de realizar predicciones precisas de la demanda energética en la planificación, generación y distribución de energía, específicamente en el contexto del departamento de Santander. Para lograr esto, se innovó respecto a los trabajos previos en este contexto, implementando modelos que constaban de tres fases esenciales: **extracción de características de los datos de entrada**, etapa que permitió ampliar el conjunto de información disponible y extraer los datos más relevantes; **selección de predictores**, en esta fase, se eligieron cuidadosamente los predictores más importantes y aquellos que tenían una fuerte relación con la variable objetivo, que en este caso era la demanda energética; y, **predicción**, donde fueron utilizados métodos de aprendizaje automático por conjuntos basados en árboles de decisión, los cuales hasta el momento no contaban con mucha exploración para el pronóstico de demanda energética a corto plazo (STLF) en Santander.

Los resultados obtenidos con este enfoque multietapa en el contexto de Santander fueron altamente satisfactorios. Estos resultados representan una contribución significativa al desarrollo energético del departamento, ya que en su mayoría, las predicciones presentadas demostraron ser más precisas que las del operador de red y superaron el rendimiento de los algoritmos utilizados por este último.

Este estudio no solo enfatiza la importancia de una predicción precisa de la demanda energética, sino que también demuestra la efectividad de un modelo que combina la extracción de características, la selección de predictores y el aprendizaje automático basado en árboles de decisión.

Estos hallazgos tienen el potencial de tener un impacto positivo en la eficiencia y la confiabilidad del suministro de energía en el departamento de Santander.

Respecto a los pilares fundamentales que contribuyeron al éxito de los resultados obtenidos, es crucial destacar la importancia del método de Descomposición en Modos Empíricos (EMD). Esto queda claramente evidenciado, ya que la mayoría de los métodos de selección de predictores optaron principalmente por las funciones de modo intrínseco de la demanda energética para realizar el pronóstico.

Además, la relevancia de la selección de predictores es innegable. Basta con analizar cómo, a pesar de contar con un conjunto inicial de información y modelos que compartían los mismos parámetros de entrenamiento, la diferencia en el rendimiento entre un caso y otro se debió principalmente a los predictores que la máquina utilizó como entrada.

Otro factor crucial que influyó en el desempeño de los algoritmos fueron los días del calendario. Dado el patrón cíclico de la demanda energética a lo largo del día (horaria) y en el transcurso de la semana (diaria), resulta imperativo que el modelo tenga en cuenta estas características al predecir el día en cuestión. Esto se hace evidente al observar que los modelos de mejor rendimiento fueron entrenados con predictores seleccionados mediante el método de Selección Secuencial de Características hacia el Frente (SFS), el cual tuvo en cuenta en gran medida los días del calendario en comparación con otros métodos. Uno de los aspectos a mejorar en relación a los días del calendario, según lo revelado por el análisis de los resultados, sería considerar el lunes como un día especial, ya que se observó que el error promedio fue mayor en estos días en todos los modelos. Esta corrección sería deseable para futuros trabajos en este campo.

Otra mejora interesante es ampliar el conjunto de datos de entrenamiento a un periodo mínimo de un año que cuente con los datos meteorológicos para las veinticuatro horas del día como se hizo en este trabajo, pudiendo aumentar el espacio de información con predictores categóricos para clasificar el mes a predecir, así como el cuartíl del año, mejorando el rendimiento de los modelos.

Adicionalmente, se tuvieron en cuenta las limitaciones del método de pronóstico continuo utilizado para realizar las predicciones de los casos con mejor desempeño en el presente trabajo. Aunque estas limitaciones resultaron en una desviación estándar del error con un desempeño inferior al obtenido por el operador de red (OR), permitieron considerar predictores más cercanos al día objetivo a predecir, lo que facilitó la captura de tendencias semanales en el consumo energético. Esto, a su vez, posibilitó un mejor desempeño en las horas de mayor consumo energético del departamento, lo que se traduce en una mejora en términos de predicción de la demanda neta en Santander.

Se sugiere que en futuros trabajos se considere la evaluación de los modelos en condiciones de incertidumbre en sus entradas, lo que implicaría adoptar un enfoque estocástico para la predicción de la demanda en el departamento. Esto permitiría reproducir un escenario con información que no siempre es completamente precisa, lo que podría ser valioso para abordar situaciones en las que la precisión de los datos puede ser variable.

### Referencias Bibliográficas

- Agudelo, A. M. F. and Barón, C. A. V. (2022). Diseño de un algoritmo para el pronóstico de la demanda de energía eléctrica a corto plazo en el MC - Santander utilizando redes neuronales e información mutua parcial. Universidad Industrial de Santander. Tesis de pregrado.
- Anggraeni, F., Adytia, D., and Ramadhan, A. W. (2021). Forecasting of Wave Height Time Series Using AdaBoost and XGBoost, Case Study in Pangandaran, Indonesia. In *2021 International Conference on Data Science and Its Applications (ICoDSA)*, pages 97–101.
- Apolinario, G. F. D., Hong, Y.-Y., Lee, Y.-D., Jiang, J.-L., and Wang, S.-S. (2021). Comparative Studies on Use of Discrete Wavelet Transform-based Feature Extraction for Peak Load Forecasting Using LSTM. In *2021 IEEE 4th International Conference on Power and Energy Applications (ICPEA)*, pages 88–92.
- Arenas, C. Y. A. (2004). Predicción de demanda de corto plazo empleando redes neuronales. Universidad Industrial de Santander. Tesis de maestría.
- Bican, B. and Yaslan, Y. (2014). A hybrid method for time series prediction using EMD and SVR. In *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 566–569.
- Cardona, J. F. M. (2021). Pronóstico de la demanda de energía en colombia a corto plazo basado en un modelo híbrido adaptativo. Universidad Nacional de Colombia. Tesis de maestría.

- Dab, K., Agbossou, K., Cardenas, A., Dube, Y., and Kelouwani, S. (2019). Sensitivity Analysis of Exogenous Variables for Load Forecasting Using Polynomial Regression. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 2560–2565. IEEE.
- Deng, W., Zhou, S., Li, Y., Mei, Y., Ren, L., Liu, Y., Kang, T., and You, J. (2022). Medium-term rolling load forecasting based on seasonal decomposition and long short-term memory neural network. In *2022 IEEE/IAS Industrial and Commercial Power System Asia (ICPS Asia)*, pages 1627–1632.
- Fan, J. and Tang, Y. (2013). An EMD-SVR method for non-stationary time series prediction. In *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, pages 1765–1770. IEEE.
- Galán, E. E. S. and Díaz, A. M. R. (2021). Guía para el cálculo del pronóstico a corto plazo de la demanda de energía eléctrica: Caso santander. Universidad Industrial de Santander. Monografía.
- Gao, Y., Fang, Y., Dong, H., and Kong, Y. (2020). A Multifactorial Framework for Short-Term Load Forecasting System as Well as the Jinan’s Case Study. *IEEE Access*, 8:203086–203096.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995.

- Kadiyala, A. and Kumar, A. (2018). Applications of python to evaluate the performance of decision tree-based boosting algorithms. *Environmental Progress & Sustainable Energy*, 37(2):618–623.
- Kumari, N., Namrata, K., Kumar, M., and Gupta, R. P. (2022). Random Forest Algorithm for Solar Forecasting in Jamshedpur – India. In *2022 4th International Conference on Energy, Power and Environment (ICEPE)*, pages 1–6.
- Laszuk, D. (2017). Python implementation of Empirical Mode Decomposition algorithm. <https://github.com/laszukdawid/PyEMD>.
- Li, C. (2020). Designing a short-term load forecasting model in the urban smart grid system. *Applied Energy*, 266:114850.
- Li, K., Xing, Y., Zhu, H., and Nai, W. (2020). A Short-Term Hybrid Forecasting Approach for Regional Electricity Consumption Based on Grey Theory and Random Forest. In *2020 5th International Conference on Computational Intelligence and Applications (ICCI)*, pages 194–198.
- Moral-Carcedo, J. and Pérez-García, J. (2017). Integrating long-term economic scenarios into peak load forecasting: An application to Spain. *Energy*, 140:682–695.
- Moral-Carcedo, J. and Pérez-García, J. (2019). Time of day effects of temperature and daylight on short term electricity load. *Energy*, 174:169–183.
- Paliari, I., Karanikola, A., and Kotsiantis, S. (2021). A comparison of the optimized LSTM, XG-

- BOOST and ARIMA in Time Series forecasting. In *2021 12th International Conference on Information, Intelligence, Systems Applications (IISA)*, pages 1–7.
- Perez, D. F. M. and Tarazona, D. J. B. (2021). Pronostico de demanda a corto plazo en sistemas eléctricos: Un caso de estudio utilizando una técnica en máquinas de aprendizaje. Universidad Industrial de Santander. Tesis de pregrado.
- Phan, Q.-T., Wu, Y.-K., and Phan, Q.-D. (2020). A Comparative Analysis of XGBoost and Temporal Convolutional Network Models for Wind Power Forecasting. In *2020 International Symposium on Computer, Consumer and Control (IS3C)*, pages 416–419.
- Prashanthi, P. and Priyadarsini., K. (2021). A Comparative Study of the Performance of Machine Learning based Load Forecasting Methods. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 132–136.
- Sweetlin, E. J. and Saudia, S. (2023). Ensemble Learning for the Survivability Prediction of Breast Cancer Patients Using METABRIC and SEER Datasets. In *2023 4th International Conference on Signal Processing and Communication (ICSPC)*, pages 16–21.
- Tajeuna, E. G., Bouguessa, M., and Wang, S. (2018). A network-based approach to enhance electricity load forecasting. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 266–275.
- Taylor, J. W., de Menezes, L. M., and McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*,

22(1):1–16.

Villamil, Y. J. D. (2016). Predicción de demanda de corto plazo aplicado a las condiciones de Colombia. Pontificia Universidad Javeriana. Tesis de maestría.

Wang, C., Zhang, J., Tian, L., Xue, L., Zheng, Y., and Liu, L. (2021). Short-term Load Forecasting Based on Kprototypes Clustering and Random Forest. In *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pages 1226–1230.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons.

WH, B. (1921). Weather and harvest cycles. *The Economic Journal*, pages 429–452.

Wu, Z., Feng, J., Qiao, F., and Tan, Z.-M. (2016). Fast multidimensional ensemble empirical mode decomposition for the analysis of big spatio-temporal datasets. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150197.

Yu, C.-J., He, Y.-Y., and Quan, T.-F. (2008). Frequency spectrum prediction method based on EMD and SVR. In *2008 Eighth International Conference on Intelligent Systems Design and Applications*, volume 3, pages 39–44. IEEE.

Zhang, W., Yan, G., and Mu, G. (2017). A novel space-time feature extraction approach for load forecasting. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 1953–1957.

Zheng, H., Yuan, J., and Chen, L. (2017). Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies*, 10(8).

Ziel, F. (2018). Modeling public holidays in load forecasting: a German case study. *Journal of Modern Power Systems and Clean Energy*, 6(2):191–207.

Zor, K., Timur, O., and Teke, A. (2017). A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. In *2017 6th International Youth Conference on Energy (IYCE)*, pages 1–7.

## Apéndices

### Apéndice A. Repositorio GitLab CPDEAM222

En el siguiente enlace puede consultar de manera libre el código utilizado para el desarrollo de este proyecto el cual corresponde a un repositorio público de GitLab.

<https://gitlab.com/sergioalug/cpdeam222>

Este repositorio cuenta con una pequeña sección instructiva para su entendimiento, uso y estudio (*Readme*).

**Apéndice B. Desviaciones calculadas bajo el Acuerdo CNO 1303 de 2020 con los resultados de los modelos XGBoost-SFS, RF-SFS y el Operador de Red (OR)**

<b>Fecha</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-07-01	0.031748	0.309734	1.101315
2023-07-02	3.478496	3.796538	3.640610
2023-07-01	0.031748	0.309734	1.101315
2023-07-02	3.478496	3.796538	3.640610
2023-07-03	0.218915	0.685245	0.313367
2023-07-04	0.673681	0.237798	3.404242
2023-07-05	2.489270	1.591752	0.432080
2023-07-06	0.875466	1.185062	1.262598
2023-07-07	0.995867	0.565577	0.687791
2023-07-08	1.012584	0.474449	0.830962
2023-07-09	0.752376	1.150196	2.518608
2023-07-10	1.074238	0.984481	0.426201
2023-07-11	1.577239	0.676372	0.773270
2023-07-12	1.015098	1.149377	2.917602
2023-07-13	0.475258	0.042097	0.946794
2023-07-14	0.384659	0.274489	0.548070
2023-07-15	0.648297	0.784947	0.809779
2023-07-16	1.069720	0.290579	2.237950
2023-07-17	1.469127	2.040322	3.004682
2023-07-18	4.530424	4.818714	5.302516
2023-07-19	2.591594	2.499316	1.733520
2023-07-20	10.323058	9.527004	8.955501
2023-07-21	1.211759	1.898605	2.179167
2023-07-22	1.123275	1.519017	3.077393
2023-07-23	2.217988	2.071704	5.200432

<b>Fecha</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-07-24	2.148663	2.595793	6.474695
2023-07-25	2.906926	3.706253	6.631421
2023-07-26	1.880008	2.286033	4.920311
2023-07-27	0.000446	0.421459	1.489683
2023-07-28	0.529613	0.013764	0.450152
2023-07-29	1.393466	2.107940	2.193521
2023-07-30	2.257791	2.723891	7.084029
2023-07-31	0.358539	0.634024	2.928777
2023-08-01	0.935808	2.063544	2.774512
2023-08-02	0.709057	0.346441	0.264472
2023-08-03	0.645352	0.490375	0.846084
2023-08-04	1.483839	0.691975	1.401955
2023-08-05	1.035712	0.413613	0.649428
2023-08-06	0.479165	0.232450	4.835254
2023-08-07	15.251419	14.405856	4.211397
2023-08-08	0.654806	1.576975	1.283162
2023-08-09	3.242021	3.639249	4.323279
2023-08-10	2.507422	2.296921	2.637246
2023-08-11	3.997236	4.255143	5.846210
2023-08-12	4.058366	4.217020	6.620458
2023-08-13	5.072472	5.430059	8.534557
2023-08-14	2.744845	2.540565	3.460452
2023-08-15	1.941013	2.209916	5.077860
2023-08-16	3.451973	3.119491	2.473613
2023-08-17	3.497363	3.533922	4.541099
2023-08-18	0.924498	0.862083	2.010272
2023-08-19	0.875203	1.680865	3.529990
2023-08-20	3.435706	3.039742	1.163132

---

<b>Fecha</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-08-21	3.363205	4.750204	2.773400
2023-08-22	0.987314	0.784074	6.053663
2023-08-23	2.001396	1.012463	2.283353
2023-08-24	2.017019	2.341345	0.472185
2023-08-25	3.155530	3.016628	1.750587
2023-08-26	1.848973	2.479878	1.882629
2023-08-27	2.274249	3.700830	7.297622
2023-08-28	1.801468	2.053631	2.107836
2023-08-29	2.237054	3.261663	4.604083
2023-08-30	3.162794	3.987970	5.662338
2023-08-31	3.565243	4.387067	6.630973

---

**Apéndice C. Desviaciones calculadas bajo la Resolución CREG 100 de 2019 con los resultados de los modelos XGBoost-SFS, RF-SFS y el Operador de Red (OR)**

<b>Fecha</b>	<b>Tramo</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-07-01	1	1.562241	1.115065	2.175884
2023-07-01	2	1.293688	0.387699	0.170719
2023-07-02	1	5.653881	5.993995	5.020248
2023-07-02	2	1.679784	1.979575	2.499858
2023-07-03	1	0.663805	0.142153	1.792881
2023-07-03	2	0.133191	1.115072	0.857588
2023-07-04	1	1.839635	1.061805	4.079671
2023-07-04	2	2.660492	1.265151	2.870307
2023-07-05	1	1.946775	1.049803	0.126783
2023-07-05	2	2.928572	2.030611	0.679303
2023-07-06	1	3.661397	3.912906	2.615178
2023-07-06	2	1.313946	0.958701	0.199630
2023-07-07	1	0.763072	1.187615	0.871840
2023-07-07	2	2.406853	1.971954	0.540150
2023-07-08	1	1.400291	0.119304	0.129365
2023-07-08	2	0.678919	0.780091	1.434764
2023-07-09	1	1.252901	0.554651	1.540667
2023-07-09	2	2.403371	2.553840	3.323771
2023-07-10	1	3.207646	2.042477	1.667440
2023-07-10	2	0.653553	0.127638	0.579045
2023-07-11	1	0.606573	0.429972	2.483393
2023-07-11	2	2.401154	0.885519	3.537576
2023-07-12	1	3.225389	3.901518	4.984090
2023-07-12	2	0.703347	0.990343	1.310960
2023-07-13	1	1.937337	1.876590	1.101859

<b>Fecha</b>	<b>Tramo</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-07-13	2	0.661736	1.384507	2.539940
2023-07-14	1	1.849175	1.726719	4.080172
2023-07-14	2	0.882984	0.982518	2.509214
2023-07-15	1	0.815942	1.407625	0.814136
2023-07-15	2	1.855960	2.593317	2.149138
2023-07-16	1	0.732333	0.186140	2.593528
2023-07-16	2	1.354556	0.693044	1.937758
2023-07-17	1	0.666668	0.083104	2.613495
2023-07-17	2	3.147340	3.708814	3.312058
2023-07-18	1	3.916417	3.924351	5.352607
2023-07-18	2	5.023952	5.537587	5.262255
2023-07-19	1	5.025591	3.371702	2.557253
2023-07-19	2	0.597849	1.784724	1.058780
2023-07-20	1	8.791035	9.356942	8.992129
2023-07-20	2	11.579107	9.666431	8.925470
2023-07-21	1	0.532270	0.205084	2.033320
2023-07-21	2	2.618536	3.264641	2.296811
2023-07-22	1	2.008523	1.348224	4.059476
2023-07-22	2	0.358365	1.666594	2.228811
2023-07-23	1	0.401290	0.632713	5.131370
2023-07-23	2	3.740269	3.277490	5.258301
2023-07-24	1	0.337328	1.094102	5.223553
2023-07-24	2	3.547899	3.755833	7.441188
2023-07-25	1	2.121627	3.794383	8.593829
2023-07-25	2	3.559751	3.632990	5.000056
2023-07-26	1	0.647490	1.378911	5.096063
2023-07-26	2	2.875489	3.018697	4.778360
2023-07-27	1	1.282276	1.034590	2.692322

<b>Fecha</b>	<b>Tramo</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-07-27	2	1.056450	1.620153	0.499609
2023-07-28	1	1.127394	1.057897	2.141636
2023-07-28	2	0.031252	0.907192	0.960012
2023-07-29	1	0.469264	0.492857	0.041998
2023-07-29	2	2.914525	3.426776	3.950401
2023-07-30	1	2.565500	4.076132	9.583823
2023-07-30	2	1.986895	1.533424	4.883297
2023-07-31	1	3.238911	4.542460	4.768359
2023-07-31	2	1.846406	2.357913	1.520563
2023-08-01	1	1.512681	2.379128	4.859164
2023-08-01	2	0.455991	1.801056	1.040594
2023-08-02	1	1.213470	0.054272	0.732157
2023-08-02	2	0.299240	0.672005	0.115505
2023-08-03	1	1.205345	0.840001	0.754799
2023-08-03	2	0.193919	0.208526	0.919673
2023-08-04	1	2.438480	0.771948	1.023450
2023-08-04	2	0.706696	0.626872	1.710084
2023-08-05	1	1.976707	1.509898	0.130262
2023-08-05	2	0.245185	0.507373	1.085578
2023-08-06	1	0.334403	0.821084	6.273188
2023-08-06	2	0.605277	0.280346	3.582581
2023-08-07	1	12.859613	12.725355	3.551478
2023-08-07	2	17.218903	15.788224	4.754243
2023-08-08	1	0.305026	0.061487	0.099276
2023-08-08	2	1.406701	2.860481	2.210570
2023-08-09	1	2.573331	2.654442	4.286708
2023-08-09	2	3.780414	4.432161	4.352725
2023-08-10	1	1.403846	1.279952	2.771954

<b>Fecha</b>	<b>Tramo</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-08-10	2	3.398724	3.118275	2.528449
2023-08-11	1	2.662878	2.944353	5.125270
2023-08-11	2	5.061167	5.300281	6.421040
2023-08-12	1	4.452768	4.560120	7.059436
2023-08-12	2	3.720699	3.923276	6.244628
2023-08-13	1	4.426559	4.312624	9.056218
2023-08-13	2	5.618602	6.374867	8.093485
2023-08-14	1	3.086717	3.123724	3.716349
2023-08-14	2	2.473625	2.077921	3.257439
2023-08-15	1	1.880020	2.722851	5.995831
2023-08-15	2	1.990830	1.790958	4.328075
2023-08-16	1	2.835595	2.578309	3.177734
2023-08-16	2	3.955508	3.561597	1.898397
2023-08-17	1	2.437103	2.508048	4.188684
2023-08-17	2	4.346748	4.355760	4.823421
2023-08-18	1	0.777543	1.319333	2.547633
2023-08-18	2	1.044665	0.488182	1.570862
2023-08-19	1	1.098841	1.931422	4.167824
2023-08-19	2	0.683313	1.465877	2.982703
2023-08-20	1	4.279874	3.927285	0.687501
2023-08-20	2	2.715030	2.282036	1.569184
2023-08-21	1	1.049771	0.390511	3.546198
2023-08-21	2	5.268227	8.340241	2.137031
2023-08-22	1	1.037216	4.729640	9.101881
2023-08-22	2	2.529559	2.221576	3.731593
2023-08-23	1	1.177723	0.125947	2.889716
2023-08-23	2	2.657741	1.919607	1.800172
2023-08-24	1	1.500153	1.453519	0.316648

<b>Fecha</b>	<b>Tramo</b>	<b>XGB-SFS</b>	<b>RF-SFS</b>	<b>OR</b>
2023-08-24	2	2.428286	3.047783	1.099856
2023-08-25	1	2.873187	3.176684	2.796144
2023-08-25	2	3.388977	2.884290	0.886098
2023-08-26	1	0.018224	0.680408	0.459616
2023-08-26	2	3.359045	3.964149	3.056385
2023-08-27	1	0.534213	1.490439	6.349661
2023-08-27	2	4.577833	5.513860	8.075168
2023-08-28	1	1.162073	1.128695	2.794197
2023-08-28	2	2.304535	2.781358	1.567818
2023-08-29	1	2.149473	2.898640	4.333518
2023-08-29	2	2.307667	3.554355	4.822229
2023-08-30	1	2.897961	2.549877	4.758252
2023-08-30	2	3.374033	5.135039	6.383465
2023-08-31	1	3.605377	4.101654	7.100620
2023-08-31	2	3.532637	4.618948	6.249411