

# Anticipando el Futuro Energético: Predicción del Consumo Eléctrico Domiciliario mediante Series Temporales

Marco Mongi<sup>1,2\*</sup>

<sup>1</sup>Ing. en Telecomunicaciones, UNRC

<sup>2</sup>Product Owner, Ascentio Technologies

## Objetivos

*El objetivo principal de este estudio es desarrollar un modelo predictivo para estimar el consumo futuro de energía eléctrica en hogares, basado en datos históricos. Para ello, se emplearán técnicas de regresión y análisis de series temporales.*

## Descripción de la problemática

El crecimiento acelerado del consumo de energía eléctrica supera actualmente la capacidad de generación, una tendencia intensificada por el aumento de la demanda de las granjas de servidores dedicadas a modelos de inteligencia artificial. Esta discrepancia entre oferta y demanda no solo ejerce presión sobre la infraestructura energética existente, sino que también incrementa los costos operativos y el riesgo de interrupciones en el suministro.

Enfrentar este desafío global requiere el desarrollo de estrategias inteligentes para equilibrar la distribución de la carga eléctrica. Una solución prometedora es la implementación de "baterías virtuales" de electricidad, que consiste en incentivar a los consumidores a desplazar su uso de energía a horarios de menor demanda. Esto optimiza la distribución energética y ofrece beneficios económicos tanto para los usuarios, que reciben incentivos, como para las compañías eléctricas, que reducen costos asociados al mantenimiento y gestión de picos de consumo.

En este contexto, la capacidad de predecir con precisión el consumo eléctrico a nivel domiciliario es de vital importancia. Comprender y anticipar los patrones de consumo permite a las empresas proveedoras tomar decisiones informadas para gestionar sus recursos de manera más eficiente, contribuyendo así a la estabilidad del sistema eléctrico y al desarrollo sostenible.

## Fuente de datos

Los datos utilizados provienen del dataset "Individual household electric power consumption" disponible en el repositorio de la Universidad de Cal-

ifornia en Irvine [1]. Este dataset contiene 2.075.259 mediciones registradas en una vivienda unifamiliar en Francia entre diciembre de 2006 y noviembre de 2010 (47 meses), con una frecuencia de un registro por minuto.

Descripción de los campos en el dataset:

- Date:** Fecha en formato dd/mm/yyyy
- Time:** Hora en formato hh:mm:ss
- global\_active\_power:** Potencia activa global promedio por minuto (kW).
- global\_reactive\_power:** Potencia reactiva global promedio por minuto (kW).
- voltage:** Voltaje promedio por minuto (V).
- global\_intensity:** Intensidad de corriente global promedio por minuto (A).
- sub\_metering\_1:** Medición de energía de zona No. 1 (Wh). Corresponde a la cocina (lavavajillas, horno, microondas).
- sub\_metering\_2:** Medición de energía de zona No. 2 (Wh). Corresponde al lavadero (lavadora, secadora, refrigerador, iluminación).
- sub\_metering\_3:** Medición de energía de zona No. 3 (Wh). Corresponde al calentador de agua eléctrico y aire acondicionado.

## Metodología

El desarrollo del modelo predictivo se llevó a cabo siguiendo estos pasos:

- Análisis Exploratorio de Datos (EDA):** Se examinó el conjunto de datos para identificar patrones, tendencias y relaciones entre variables.
- Preprocesamiento del Dataset:** Incluyó limpieza de datos, transformación de variables y creación de nuevas características.
- Selección y Entrenamiento de Modelos:** Se probaron modelos Random Forests, redes neu-

\*mail: marcomongi@gmail.com

Sent: October 20, 2024

ronales Feed Forward y Prophet para la predicción de series temporales.

4. **Evaluación y Comparación de Modelos:** Se utilizaron métricas de error para determinar el modelo más preciso y eficiente.

El desarrollo de este trabajo se realizó teniendo en cuenta los fundamentos aprendidos a lo largo de la cursada, según los materiales de clase [2].

## Análisis Exploratorio de Datos (EDA)

Se realizó un análisis detallado para comprender las características del consumo eléctrico:

**Distribución del Consumo:** El histograma de la Figura 1 reveló una distribución bimodal, confirmada con la prueba de Hartigan's Dip Test. El consumo medio de potencia aparente es de 1,11 kW, con valores mínimos y máximos de 0 kW y 11,1 kW, respectivamente. La desviación estándar es de 1,05 kW.

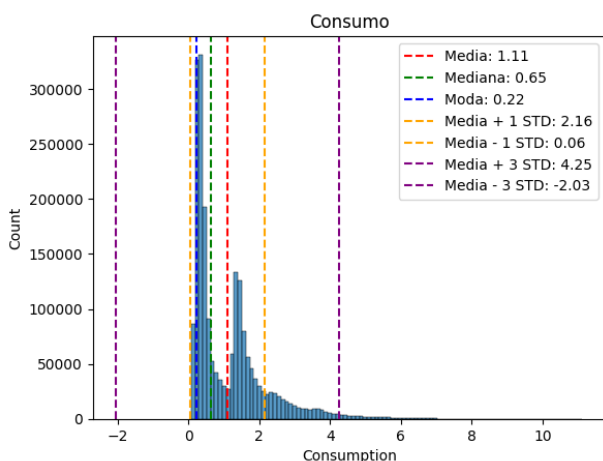


Figure 1: Histograma valores de potencia aparente.

**Consumo Activo vs. Reactivo:** El consumo activo representa el 89,8% del total, mientras que el reactivo es el 10,2%. La zona de climatización muestra el mayor consumo energético, según la Figura 2.

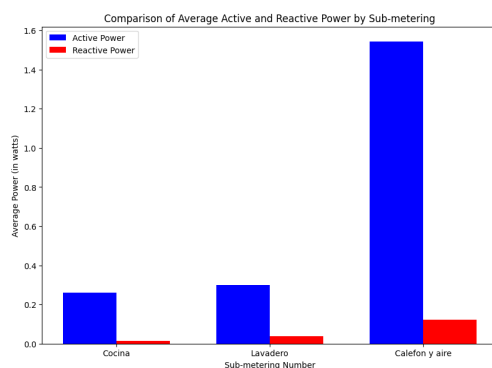


Figure 2: Potencia activa y reactiva para las distintas zonas

**Patrones Temporales diarios:** Aumentos en el consumo durante las horas de despertar y regreso del trabajo; disminución durante la noche (Figura 3).

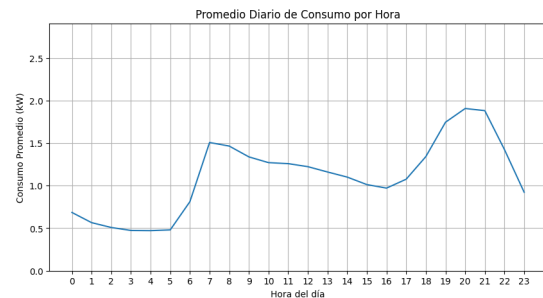


Figure 3: Consumo de potencia según la hora del día

**Patrones Temporales anuales:** Patrón en forma de "V" con consumo mínimo en agosto (verano) y picos en diciembre y enero (invierno), indicando una relación inversa entre temperatura y consumo (Figura 4).

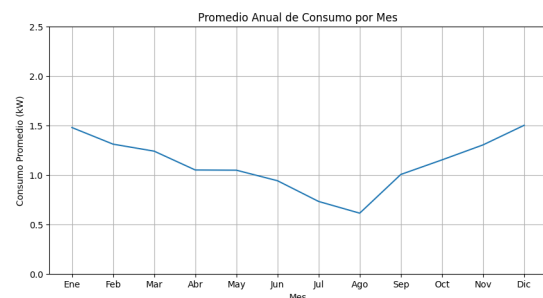


Figure 4: Consumo de potencia según el mes del año

Analizando la temperatura promedio a lo largo del año en Francia, se observa que los picos de consumo coinciden con las temperaturas más frías (Figura 5).

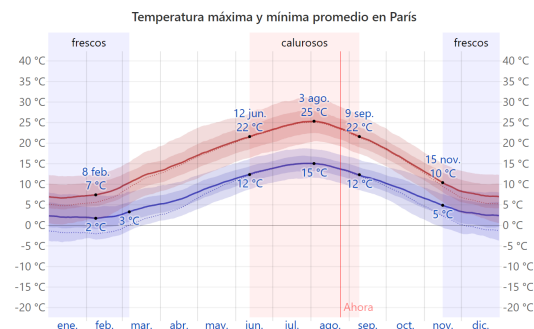


Figure 5: Temperatura ambiente según el mes del año

Por otra parte, el consumo de los dispositivos de calefacción es el mayor (Figura 6). Con esta conclusión y la anterior puede intuirse que los picos de consumo entre diciembre y enero se deben a que en esta época debido a las bajas temperaturas, se encienden los calefactores eléctricos.

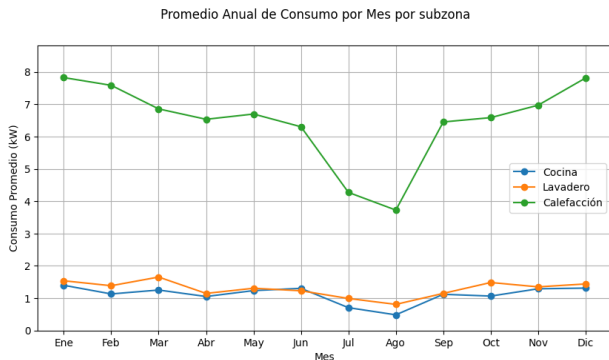


Figure 6: Consumo de potencia según el mes del año por zona

**Análisis de Frecuencia:** La Transformada de Fourier identificó periodicidades significativas cada medio día, un día y un año (Figura 7.), validando las ventanas temporales seleccionadas.

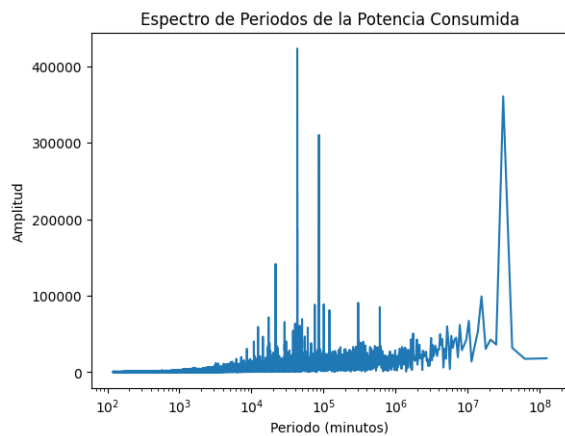


Figure 7: Potencia consumida en dominio de Fourier

Análisis adicionales incluyen:

- **Consumo Año a Año:** Patrones consistentes sin variaciones significativas.
- **Correlación entre Variables:** Relaciones lineales claras, especialmente entre potencia aparente y activa, e intensidad de corriente. Se evidencian mediante pairplot y heatmap.
- **Voltaje y Corriente en el Tiempo:** Fluctuaciones mínimas, indicando estabilidad en el suministro eléctrico.
- **Consumo por Día de la Semana y Día del Mes:** Sin variaciones apreciables, sugiriendo independencia del consumo respecto a estas variables.

Estos análisis, así como el código que respalda las conclusiones anteriores y un análisis en más detalle, puede consultarse en el Notebook asociado al EDA presentado junto con este trabajo [3].

## Preprocesamiento del dataset

Se implementaron las siguientes acciones:

- **Limpieza de Datos:** Reemplazo de caracteres "?" por valores faltantes y posterior interpolación lineal para completar datos ausentes.
- **Conversión de Formatos:** Ajuste de tipos de datos adecuados (fechas, enteros, flotantes).
- **Ingeniería de Características:**

- Unificación de fecha y hora en una sola variable datetime.
- Cálculo de la potencia aparente (S) como la raíz cuadrada de la suma de los cuadrados de las potencias activa (P) y reactiva (Q), según Ecuación 1.

$$S = \sqrt{P^2 + Q^2} \quad (1)$$

- Eliminación de variables altamente correlacionadas (intensidad de corriente y potencia activa) según test de Pearson para simplificar el modelo según principio de parsimonia.
- Descarte de variables con poca variabilidad (voltaje) y de submediciones para mejorar la generalización (se desea predecir sólo en base a mediciones anteriores).
- Creación de variables temporales: extracción de hora, día de la semana y mes.
- Estandarización de datos para modelos sensibles a escalas (redes neuronales).
- Inclusión del consumo previo como predictor en redes neuronales.
- **Tratamiento de Outliers:** No se identificaron valores atípicos que requirieran intervención. Los valores del dataset estaban dentro de rangos de consumo adecuados para una vivienda unifamiliar.
- **Resampling Temporal:** Reducción de la granularidad de un minuto a una hora para optimizar el procesamiento y reducir ruido.

## Entrenamiento y evaluación de modelos

Se abordó la predicción como un problema de series temporales sin variables exógenas. Se evalúan los siguientes modelos:

### 1. Random Forests:

- **Predictores:** 'hour', 'day\_of\_week', 'month'.
- **Parámetros del modelo:** 100 árboles.

### 2. Redes Neuronales Feedforward:

- **Arquitectura:** 3 capas ocultas (128, 64, 32 neuronas) con función de activación ReLU. 1 neurona en la capa de salida.
- **Predictores:** Mismos que Random Forests, más el consumo previo. Se aplicó un escalador estándar a los datos de los predictores.

### 3. Prophet:

- **Modelo:** Descompone la serie en tendencia, estacionalidad y efectos de eventos. Es recomendable para series temporales con comportamientos periódicos. Se puede consultar en mucho mayor detalle el funcionamiento de este algoritmo en el Notebook desarrollado para este modelo [4].
- **Predictores:** fecha y hora

El análisis en Notebook de Random Forests corresponde a [5], el de Redes Neuronales a [6], y el de Prophet a [4].

Para el modelo Random Forests, se deja como recurso en [7] una interfaz gráfica interactiva (desarrollada con Gradio) que permite hacer predicciones entre fechas a demanda, y otra que permite comparar predicciones contra valores reales para el rango de fechas del dataset.

Se presentan a continuación, los parámetros generales relativos al entrenamiento:

- **Periodo de Entrenamiento:** 16/12/2006 17:00 al 01/01/2010 00:00 (77% de los datos).
- **Periodo de Prueba:** 01/01/2010 00:00 al 26/11/2010 21:00 (22% de los datos)
- **Consideraciones:** No se mezclaron los datos (shuffling) ni se utilizaron conjuntos de validación debido a su naturaleza temporal.

### Comparación de resultados

Se evaluaron los modelos utilizando MSE, MAE y MAPE para el dataset completo. Los resultados se presentan en la Figura 8.

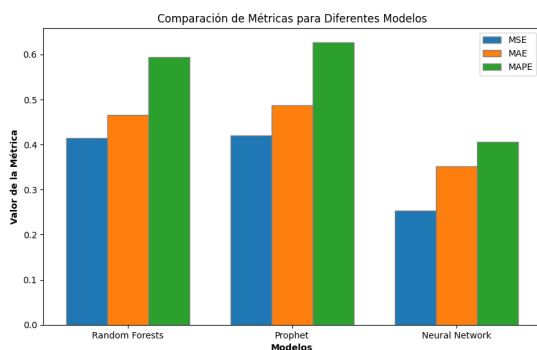


Figure 8: Métricas MSE, MAE y MAPE para el dataset de test

Se evidencia de los resultados obtenidos:

- **Redes Neuronales:** Mejores métricas, pero su dependencia del consumo previo limita su aplicabilidad para predicciones a largo plazo sin un pipeline de retroalimentación.

- **Prophet:** Rendimiento consistente y capacidad para modelar estacionalidades y tendencias, superando ligeramente a Random Forests.
- **Random Forests:** Resultados aceptables pero menos precisos en comparación.

Dado que el rendimiento de los modelos predictores de series temporales depende del tamaño del horizonte a predecir, se hizo un análisis independiente para distintos horizontes: 1 día, 1 semana, 1 mes, 3 meses, 6 meses y 1 año, analizando la MAE para cada período. El análisis por ventanas temporales mostró que, de forma poco intuitiva, el error es mayor en horizontes más cortos, estabilizándose a partir de una semana. Las gráficas de error en la predicción de la Figura 9 confirmaron que todos los modelos capturan adecuadamente los patrones de consumo.

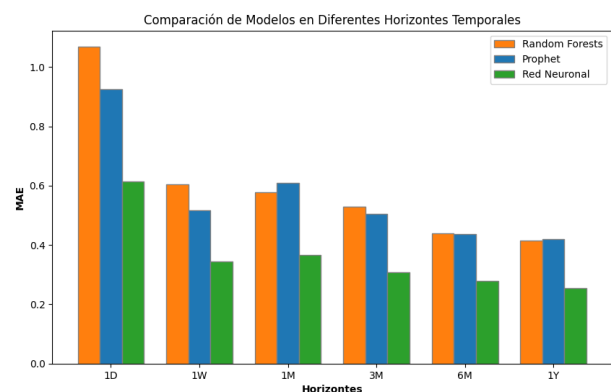
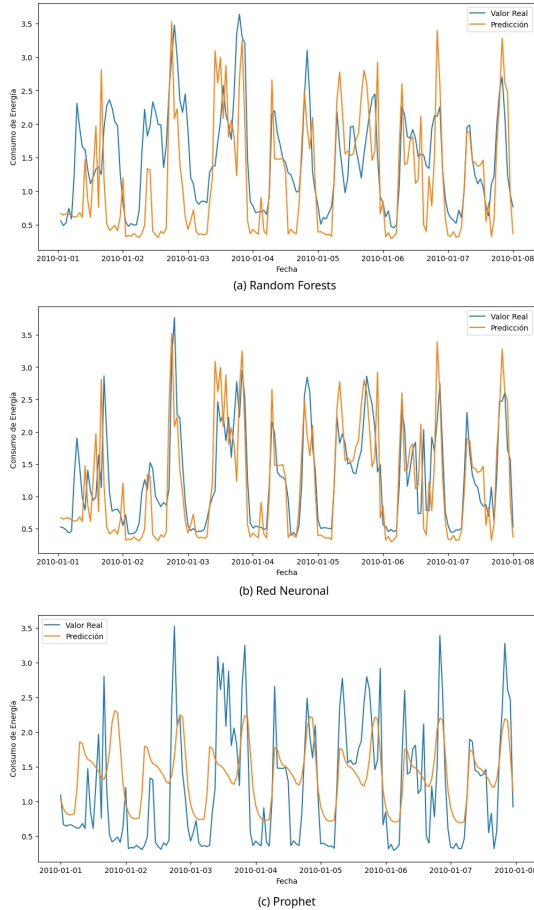


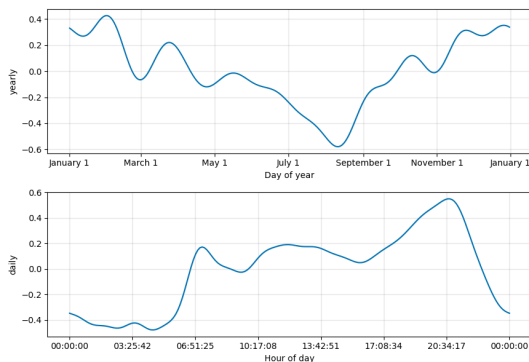
Figure 9: MSE para diferentes ventanas temporales en el dataset de test

A continuación, se presentan las gráficas de las predicciones versus los valores reales para los distintos modelos, a fin de presentar visualmente qué tan bien ajustan los modelos a los datos. Se elige una ventana de 1 semana por interpretabilidad visual. En la Figura 10 se muestra la gráfica para Random Forests (a), para la Red Neuronal (b), y para Prophet (c).



**Figure 10:** Predicción de una semana para (a) Random Forests, (b) Redes Neuronales, (c) Prophet

Un aspecto interesante a destacar de Prophet es que en su descomposición de la serie temporal en componentes de tendencia, periódicas, de feriados y residuales, las gráficas de las series de Fourier que modelan las componentes periódicas configuradas cada 1 día y anual según la Figura 11 son acordes a las Figuras 3 y 5 de los consumo promedio en distintas ventanas temporales obtenidas en el EDA.



**Figure 11:** Descomposición de la función a modelar en componentes de Prophet

## Resultados

El modelo Prophet se identifica como la opción más adecuada para predecir el consumo futuro de energía eléctrica en hogares.

Aunque las redes neuronales mostraron métricas superiores, el hecho de depender del valor previo de consumo implica que sólo se podría predecir el valor de consumo inmediatamente posterior al último entrenado, y para valores futuros es necesario realimentar este valor predicho a la red, e iterar hasta llegar al instante a predecir. De esta forma se propaga el error al aplicar este pipeline de backtesting. Dado que en este trabajo no se aplicó dicho pipeline para realizar la predicción, los resultados obtenidos para redes neuronales no pueden ser comparados contra los de Prophet y Random Forests.

Prophet ofrece un equilibrio entre precisión y simplicidad, aprovechando su diseño para series temporales con componentes estacionales. Su capacidad para descomponer la serie en tendencia y estacionalidad proporciona insights valiosos para las compañías eléctricas en la planificación y gestión de recursos.

### Futuras Mejoras:

- Incorporar variables exógenas (e.g., condiciones meteorológicas) para enriquecer el modelo.
- Explorar modelos de redes neuronales recurrentes (e.g., LSTM) que capturen dependencias temporales más complejas.

## Bibliografía

- [1] G. Hebrail and A. Berard. *Individual Household Electric Power Consumption*. UCI Machine Learning Repository. [Dataset]. 2006. URL: <https://doi.org/10.24432/C58K54>.
- [2] M. Aguilar, F. F. Martin, and G. Rodríguez. *Materiales de cursada, Diplomatura en Data Science*. MundosE, Córdoba. 2024.
- [3] M. Mongi. *Visualización, análisis y preprocesado de datos.ipynb*. Río Cuarto. 2024.
- [4] M. Mongi. *Modelo Prophet.ipynb*. Río Cuarto. 2024.
- [5] M. Mongi. *Modelo Random Forests.ipynb*. Río Cuarto. 2024.
- [6] M. Mongi. *Modelo Feed Forward Neural Network.ipynb*. Río Cuarto. 2024.
- [7] M. Mongi. *GUI Predicciones.ipynb*. Río Cuarto. 2024.