

---

# Predicción de Demanda Energética con Inteligencia Artificial

---

Memoria del proyecto

**Autor:** Francisco Javier Gómez Pulido

Correo electrónico: [frangomezpulido2002@gmail.com](mailto:frangomezpulido2002@gmail.com)

LinkedIn: [linkedin.com/in/frangomezpulido](https://www.linkedin.com/in/frangomezpulido)

GitHub: [github.com/fragompul](https://github.com/fragompul)

## **Resumen**

Este proyecto desarrolla un modelo basado en Inteligencia Artificial para predecir la demanda energética en España. Utilizando datos históricos de consumo energético, variables meteorológicas y técnicas avanzadas de aprendizaje automático, se busca anticipar la demanda con precisión para mejorar la planificación y gestión del sistema eléctrico.

# Índice general

<b>1. Introducción</b>	<b>2</b>
1.1. Contexto y Motivación	2
1.2. Objetivos del Proyecto	2
1.3. Estructura del Documento	3
<b>2. Revisión del Estado del Arte</b>	<b>4</b>
2.1. Demanda Energética: Conceptos y Factores	4
2.2. Modelos Predictivos con IA en Energía	5
2.3. Proyectos y Estudios Relacionados	5
<b>3. Datos Utilizados</b>	<b>7</b>
3.1. Fuentes de Datos	7
3.2. Descripción de Variables	7
3.3. Preprocesamiento	8
<b>4. Metodología</b>	<b>10</b>
4.1. Planteamiento del Problema	10
4.2. Modelos Utilizados	12
4.2.1. Facebook Prophet	12
4.2.2. Random Forest Regressor	12
4.3. Validación	13
<b>5. Resultados</b>	<b>14</b>
5.1. Desempeño de los Modelos	14
5.2. Interpretación de Resultados	15
5.3. Limitaciones Observadas	16
<b>6. Implementación</b>	<b>17</b>
6.1. Arquitectura del Proyecto	17
6.2. Persistencia y Carga de Modelos	17
6.3. Carga y Preparación de Datos	18
6.4. Interfaz en Streamlit	18
6.5. Predicciones y Visualización	19
6.6. Distribución y Ejecución	19
6.7. Ventajas del Enfoque	20
<b>7. Conclusiones y Trabajo Futuro</b>	<b>21</b>
7.1. Conclusiones	21
7.2. Trabajo Futuro	22

# Capítulo 1

## Introducción

### 1.1. Contexto y Motivación

La transición energética y la creciente penetración de las energías renovables en el sistema eléctrico español están transformando de forma significativa la planificación y operación de la red. La variabilidad de la generación renovable —especialmente de fuentes como la eólica y la solar— introduce nuevos retos para equilibrar en tiempo real la oferta y la demanda eléctrica. En este contexto, disponer de modelos predictivos precisos para la demanda energética resulta esencial para garantizar la seguridad de suministro, optimizar la gestión de recursos y reducir los costes operativos [Weron \(2014\)](#).

España, como miembro de la Unión Europea, se enfrenta a compromisos de descarbonización alineados con el Pacto Verde Europeo y la Ley de Cambio Climático y Transición Energética. Esto implica una electrificación creciente de sectores como el transporte o la calefacción, lo que incrementa la complejidad de las predicciones de consumo. Además, fenómenos meteorológicos extremos y fluctuaciones económicas influyen de forma directa en los patrones de demanda, lo que refuerza la necesidad de contar con herramientas avanzadas de pronóstico [Kowalska-Pyzalska et al. \(2023\)](#).

En este escenario, la Inteligencia Artificial (IA) ofrece un conjunto de técnicas potentes para modelar la dinámica de la demanda energética, capturando tanto tendencias a largo plazo como variaciones estacionales y eventos atípicos [Wang et al. \(2019\)](#). El uso de algoritmos de aprendizaje automático y modelos específicos para series temporales, como Facebook Prophet, permite no solo mejorar la precisión de las predicciones, sino también ofrecer estimaciones con intervalos de confianza, facilitando la toma de decisiones por parte de operadores y planificadores.

### 1.2. Objetivos del Proyecto

El objetivo general de este trabajo es diseñar e implementar un sistema de predicción de la demanda energética en España basado en técnicas de Inteligencia Artificial y datos abiertos. A partir de este objetivo global, se establecen los siguientes objetivos específicos:

- **Recopilar y procesar** datos históricos relevantes de generación y demanda eléctrica en España, integrando variables energéticas, económicas y demográficas.

- **Analizar y explorar** dichos datos para identificar patrones temporales, correlaciones y tendencias de fondo.
- **Implementar y comparar** modelos de predicción, incluyendo *Facebook Prophet* y algoritmos de aprendizaje automático como *Random Forest*.
- **Evaluar el rendimiento** de los modelos mediante métricas estándar como el MAE, RMSE y MAPE, aplicando técnicas de validación temporal.
- **Desarrollar un dashboard interactivo** con *Streamlit* que permita visualizar datos históricos, predicciones y análisis comparativos de modelos.

### 1.3. Estructura del Documento

El presente documento se organiza de la siguiente manera:

- En el **Capítulo 2** se presenta una revisión del estado del arte, abordando conceptos fundamentales de la demanda energética, así como los principales enfoques basados en Inteligencia Artificial aplicados a este dominio.
- El **Capítulo 3** describe los datos utilizados, incluyendo su procedencia, variables relevantes y el proceso de preprocesamiento aplicado.
- En el **Capítulo 4** se detalla la metodología, planteando el problema como una tarea de regresión temporal y describiendo los modelos implementados.
- El **Capítulo 5** expone los resultados obtenidos, evaluando cuantitativamente el rendimiento de los modelos y discutiendo sus implicaciones.
- El **Capítulo 6** describe la implementación práctica mediante un dashboard interactivo y analiza sus posibles usos.
- Finalmente, en el **Capítulo 7** se presentan las conclusiones y se proponen líneas de trabajo futuro.

En conjunto, este trabajo busca demostrar cómo la integración de datos abiertos, técnicas avanzadas de modelado y herramientas interactivas de visualización puede contribuir a mejorar la planificación energética en un entorno cada vez más complejo y dinámico.

# Capítulo 2

## Revisión del Estado del Arte

El estudio de la demanda energética y su predicción ha sido objeto de investigación continua durante décadas. La llegada de nuevas fuentes de datos, el aumento de la capacidad de cómputo y el desarrollo de algoritmos de aprendizaje automático han impulsado avances significativos en la precisión y la flexibilidad de los modelos. En este capítulo se revisan los conceptos fundamentales, los enfoques basados en Inteligencia Artificial aplicados al ámbito energético y algunos proyectos relevantes que sirven de referencia para el presente trabajo.

### 2.1. Demanda Energética: Conceptos y Factores

La **demanda energética** se refiere a la cantidad de energía que un sistema requiere en un periodo determinado. En el caso del sistema eléctrico, se mide habitualmente en megavatios-hora (MWh) o gigavatios-hora (GWh), y está sujeta a variaciones tanto *predecibles* como *aleatorias*.

Entre los factores que influyen en la demanda destacan:

- **Factores climáticos:** la temperatura, la radiación solar, el viento o la humedad relativa afectan directamente a los patrones de consumo. Por ejemplo, las olas de calor incrementan el uso de aire acondicionado, mientras que las olas de frío intensifican el uso de calefacción [Hong et al. \(2010\)](#).
- **Factores socioeconómicos:** la actividad industrial, el crecimiento demográfico y el nivel de urbanización condicionan la demanda agregada.
- **Factores temporales:** el día de la semana, la hora del día o la estacionalidad anual generan patrones cíclicos en el consumo.
- **Eventos extraordinarios:** huelgas, emergencias sanitarias o eventos deportivos de gran magnitud pueden modificar temporalmente la demanda.

Históricamente, la predicción de demanda eléctrica se ha abordado mediante métodos estadísticos como la *regresión lineal*, los modelos *ARIMA* y variantes como *SARIMA*, que capturan estacionalidades. Sin embargo, estos enfoques presentan limitaciones para modelar relaciones no lineales o manejar múltiples variables exógenas de forma eficiente [Hippert et al. \(2001\)](#).

## 2.2. Modelos Predictivos con IA en Energía

La irrupción de la **Inteligencia Artificial** ha transformado el enfoque de la predicción energética. Los modelos basados en aprendizaje automático y *deep learning* han demostrado superar en muchos casos a los métodos estadísticos clásicos, especialmente cuando se dispone de conjuntos de datos ricos y variables heterogéneas.

Algunos enfoques relevantes incluyen:

- **Árboles de decisión y *ensemble methods*:** algoritmos como *Random Forest* o *Gradient Boosting* destacan por su capacidad para modelar relaciones no lineales, manejar interacciones complejas entre variables y ofrecer medidas de importancia de variables.
- **Modelos basados en redes neuronales:** las redes neuronales recurrentes (RNN), y especialmente las variantes *Long Short-Term Memory* (LSTM) y *Gated Recurrent Unit* (GRU), han mostrado gran rendimiento en predicciones de series temporales por su capacidad para mantener dependencias a largo plazo [Marino et al. \(2016\)](#).
- **Modelos específicos de series temporales:** *Facebook Prophet* es un modelo aditivo diseñado para capturar tendencias, estacionalidades y efectos festivos de forma interpretable, siendo especialmente útil cuando los datos presentan variaciones anuales y no linealidades suaves [Taylor and Letham \(2018\)](#).
- **Modelos híbridos:** combinaciones de modelos estadísticos y de IA, o bien de varios modelos de machine learning, para aprovechar las ventajas de cada enfoque y reducir errores residuales.

Un aspecto relevante es la incorporación de **variables exógenas** en la predicción, como indicadores meteorológicos, datos económicos o medidas de generación renovable. Estos factores pueden mejorar la capacidad predictiva, pero también introducen nuevos retos, como el alineamiento temporal de datos y la propagación de incertidumbre desde predicciones de variables intermedias.

## 2.3. Proyectos y Estudios Relacionados

Existen múltiples estudios que han abordado el problema de la predicción energética con enfoques similares al de este trabajo:

- **Predicción de demanda a corto plazo:** Hong et al. [Hong et al. \(2010\)](#) analizaron la influencia de la temperatura en la predicción de carga eléctrica diaria, incorporando modelos no lineales y obteniendo mejoras significativas frente a métodos lineales clásicos.
- **Uso de *deep learning* en energía:** Marino et al. [Marino et al. \(2016\)](#) utilizaron redes LSTM para modelar patrones complejos en series de consumo energético, demostrando su superioridad frente a modelos tradicionales en horizontes de predicción de varios días.

- **Modelos híbridos y multivariantes:** estudios recientes como el de Wang et al. [Wang et al. \(2019\)](#) recogen casos en los que la combinación de métodos —por ejemplo, un modelo Prophet para capturar tendencia y estacionalidad, junto con un Random Forest para modelar efectos no lineales— logra mejoras sustanciales en precisión.
- **Aplicaciones en España:** diversos trabajos han explorado la predicción de demanda y generación renovable en el contexto español, aprovechando datos de Red Eléctrica de España (REE) y AEMET, aunque muchos se han centrado en series univariantes. El presente trabajo busca ampliar esta perspectiva integrando múltiples fuentes y ofreciendo un despliegue visual interactivo.

En resumen, la literatura muestra que no existe un único modelo que funcione de forma óptima en todos los contextos. La elección depende de factores como la disponibilidad y calidad de datos, el horizonte de predicción, la variabilidad de la señal y las restricciones operativas. El enfoque adoptado en este proyecto —basado en datos abiertos, Prophet y Random Forest, con una plataforma interactiva para explorar resultados— se alinea con las tendencias actuales de investigación aplicada y aporta una solución adaptada al contexto español.



# Capítulo 3

## Datos Utilizados

La calidad y adecuación de los datos constituyen la base de cualquier modelo predictivo fiable. En este capítulo se describen las fuentes de datos empleadas, las variables seleccionadas para el modelado y el proceso de preprocesamiento aplicado para garantizar la coherencia y utilidad de la información.

### 3.1. Fuentes de Datos

La principal fuente de datos utilizada en este proyecto es el conjunto de datos de energía de **Our World in Data** (OWID) [Ritchie et al. \(2023\)](#). OWID recopila información de diversas agencias oficiales, como la Agencia Internacional de la Energía (IEA) y *BP Statistical Review of World Energy*, armonizando y publicando las series históricas de forma abierta y en formatos accesibles.

Este dataset contiene registros anuales para más de 200 países, abarcando desde mediados del siglo XX hasta la actualidad. Para el caso de España, incluye métricas de generación y consumo eléctrico, desagregadas por fuente de energía (renovable, fósil, nuclear, etc.), así como variables económicas y demográficas.

Además, se han utilizado datos demográficos y macroeconómicos (PIB y población) también integrados en el dataset de OWID, lo que facilita el análisis conjunto sin necesidad de combinar múltiples fuentes.

### 3.2. Descripción de Variables

A partir del conjunto original, se seleccionaron aquellas variables más relevantes para el objetivo del proyecto. La Tabla [3.1](#) resume las variables utilizadas, su descripción y unidades.

Cuadro 3.1: Variables utilizadas en el modelado

Variable	Descripción	Unidad
date	Fecha de referencia (1 de enero de cada año)	Año
electricity_generation	Generación total de electricidad	GWh
renewables_electricity	Generación eléctrica con fuentes renovables	GWh
fossil_electricity	Generación eléctrica con combustibles fósiles	GWh
coal_electricity	Generación eléctrica con carbón	GWh
gas_electricity	Generación eléctrica con gas natural	GWh
nuclear_electricity	Generación eléctrica de origen nuclear	GWh
oil_electricity	Generación eléctrica con petróleo	GWh
hydro_electricity	Generación hidroeléctrica	GWh
solar_electricity	Generación solar fotovoltaica	GWh
wind_electricity	Generación eólica	GWh
gdp	Producto Interior Bruto (PIB)	USD cte
population	Población total	Habitantes

La elección de estas variables responde a dos criterios fundamentales: su relevancia directa para explicar variaciones en la demanda/generación eléctrica y su disponibilidad completa para el periodo de análisis.

### 3.3. Preprocesamiento

El preprocesamiento de datos es un paso crucial para evitar sesgos y problemas en el entrenamiento de modelos. Las transformaciones realizadas incluyen:

1. **Filtrado geográfico:** se extrajeron exclusivamente los registros correspondientes a España (`country = "Spain"`).
2. **Conversión de fechas:** se creó una variable temporal `date` a partir del año, estableciendo como fecha de referencia el 1 de enero de cada año (`YYYY-01-01`).
3. **Selección de columnas:** se descartaron las variables irrelevantes o con un alto porcentaje de valores nulos para el caso de estudio.
4. **Tratamiento de valores faltantes:** en este dataset concreto, los valores están completos para España en el rango temporal de interés, por lo que no fue necesario aplicar imputación.
5. **Almacenamiento optimizado:** el conjunto procesado se guardó en el archivo CSV `energy_spain_clean.csv` para su uso directo en notebooks y en el dashboard.

La Figura 3.1 muestra la evolución histórica de la generación eléctrica total en España, lo que permite apreciar de forma visual tendencias y cambios estructurales, como el incremento de la generación renovable en las últimas décadas.

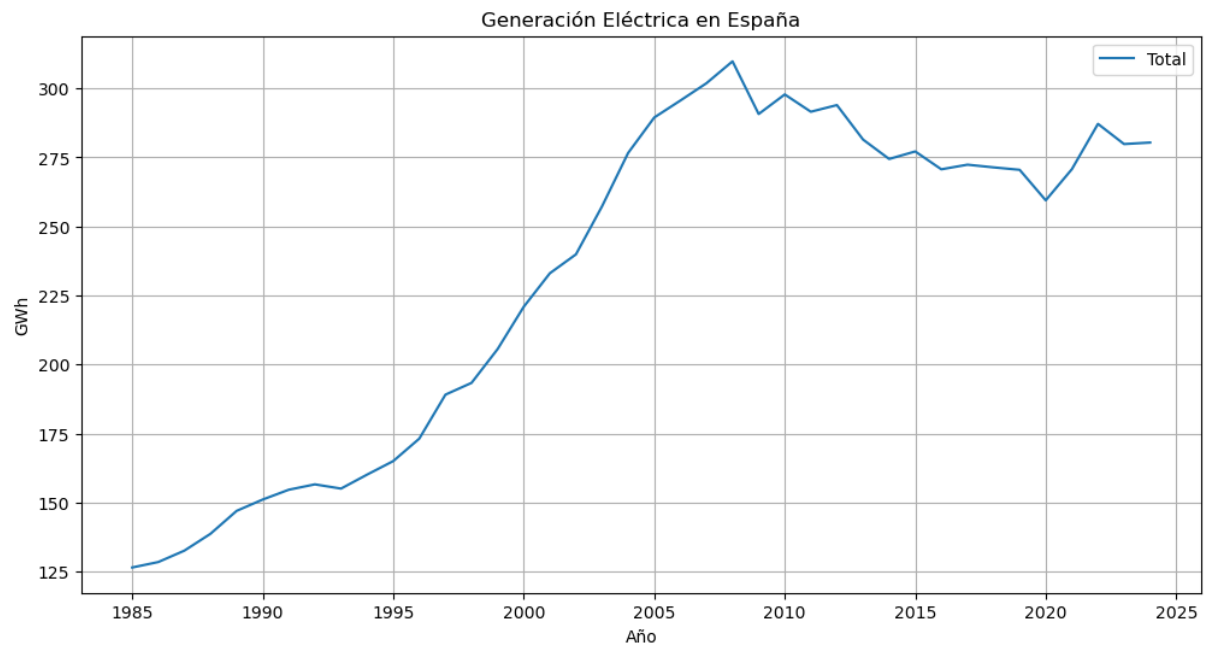


Figura 3.1: Evolución histórica de la generación eléctrica total en España (GWh).

Este conjunto de datos procesado constituye la base para el análisis exploratorio, el entrenamiento de modelos y la implementación del dashboard interactivo descrito en capítulos posteriores.

# Capítulo 4

## Metodología

La metodología seguida en este trabajo combina técnicas clásicas de análisis de series temporales con algoritmos de *machine learning* capaces de capturar relaciones no lineales entre variables. Este enfoque híbrido permite beneficiarse tanto de la interpretabilidad de modelos específicos para series temporales como de la capacidad predictiva de modelos multivariantes.

### 4.1. Planteamiento del Problema

El objetivo del proyecto se formula como un problema de **regresión multivariante de series temporales**. Dado un conjunto de variables energéticas, económicas y demográficas medidas anualmente, se busca predecir la generación total de electricidad en España para un horizonte temporal a medio plazo.

Formalmente, sea  $y_t$  la generación eléctrica total en el año  $t$  y  $\mathbf{x}_t$  el vector de variables explicativas (por ejemplo, generación renovable, generación fósil, PIB, población). El problema consiste en estimar una función:

$$\hat{y}_{t+h} = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots)$$

donde  $h$  representa el horizonte de predicción (en años). Este planteamiento permite incorporar dependencias temporales y relaciones entre múltiples variables.

El análisis exploratorio previo (Figuras 4.1 y 4.2) mostró que la generación renovable, la generación fósil y el PIB presentan correlaciones significativas con la generación total, justificando su uso como predictores clave.

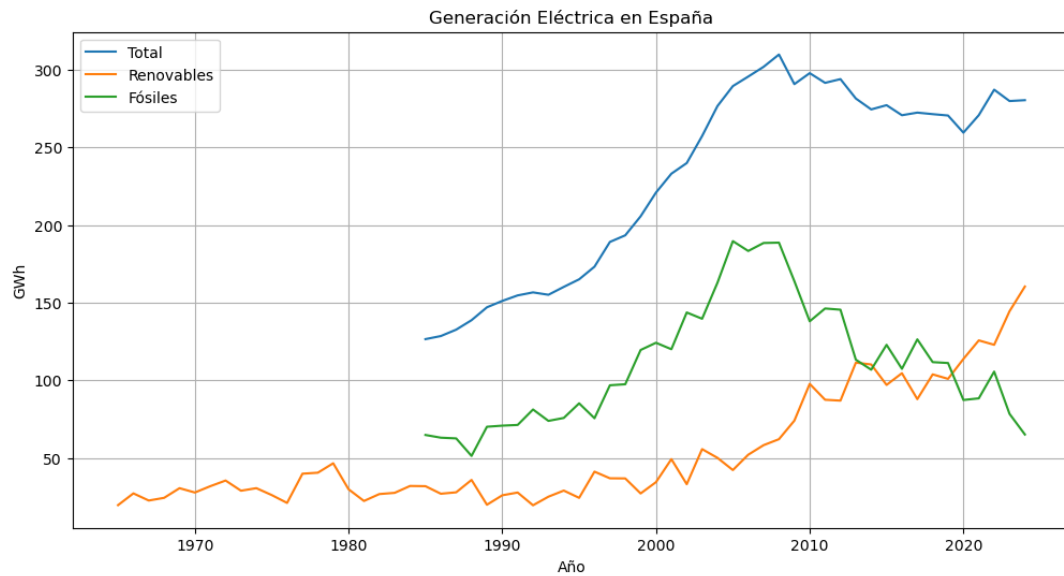


Figura 4.1: Evolución histórica de la generación eléctrica total por renovables y fósiles.

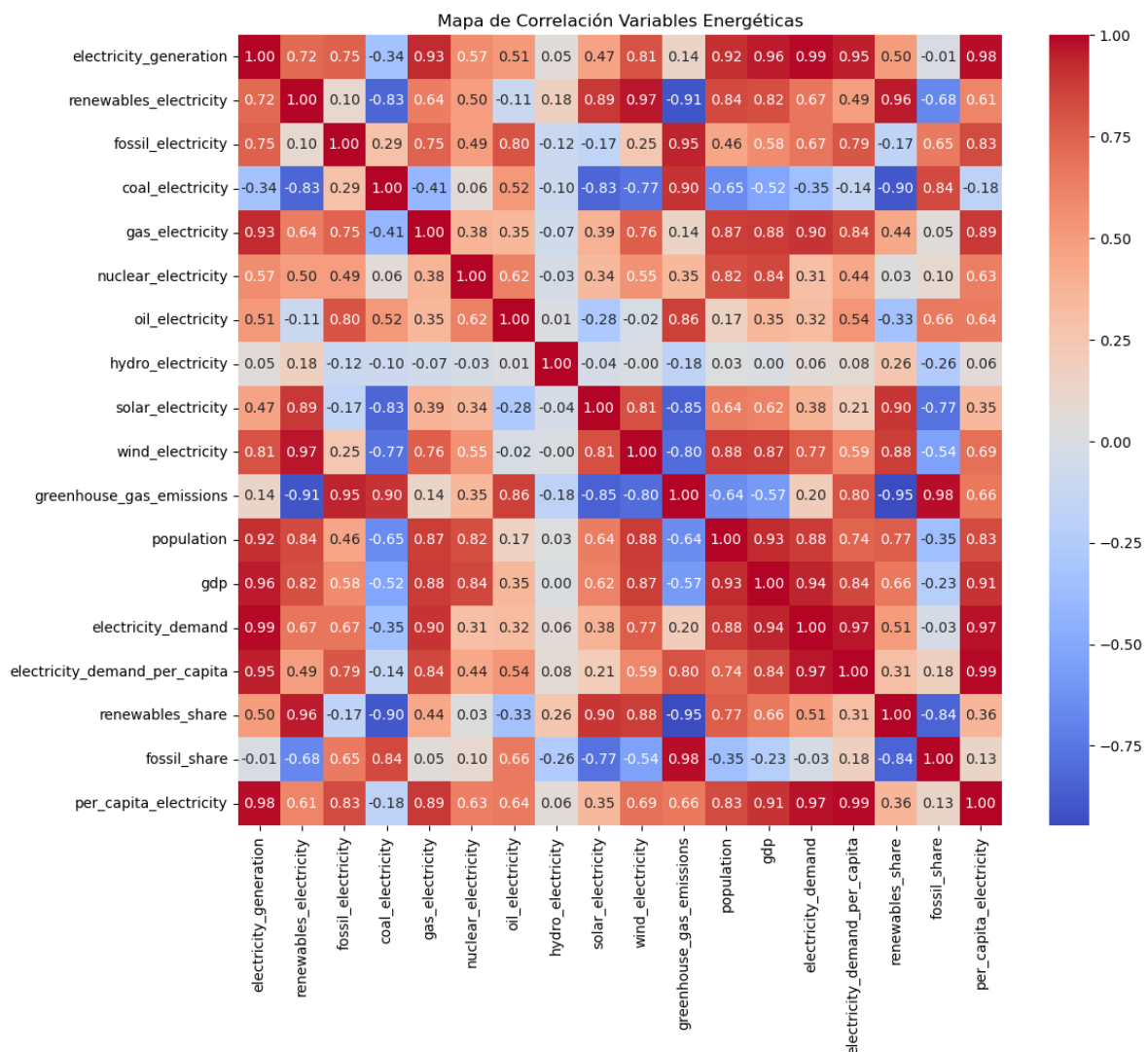


Figura 4.2: Mapa de correlación entre variables energéticas y económicas.

## 4.2. Modelos Utilizados

Se seleccionaron dos enfoques complementarios para el modelado:

### 4.2.1. Facebook Prophet

*Facebook Prophet* [Taylor and Letham \(2018\)](#) es un modelo aditivo diseñado para pronosticar series temporales con tendencias no lineales, estacionalidad múltiple y posibles cambios de tendencia. Su formulación es:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

donde:

- $g(t)$  modela la tendencia a largo plazo.
- $s(t)$  captura estacionalidades anuales, semanales u otras.
- $h(t)$  incluye efectos de festivos o eventos especiales.
- $\varepsilon_t$  es el término de error.

En este trabajo, Prophet se aplicó sobre la serie univariante de generación eléctrica total (`electricity_generation`), generando predicciones hasta 10 años en el futuro con intervalos de confianza.

### 4.2.2. Random Forest Regressor

El *Random Forest Regressor* [Breiman \(2001\)](#) es un algoritmo de *ensemble* basado en árboles de decisión que combina múltiples árboles entrenados sobre subconjuntos aleatorios de datos y características. Es capaz de modelar relaciones no lineales y manejar interacciones complejas entre variables.

Para este proyecto, el modelo se entrenó con las siguientes variables predictoras:

- Variables energéticas: `renewables_electricity`, `fossil_electricity`, `coal_electricity`, `gas_electricity`, `nuclear_electricity`, `oil_electricity`, `hydro_electricity`, `solar_electricity`, `wind_electricity`.
- Variables económicas y demográficas: `gdp`, `population`.

Este modelo ofrece, además, una estimación de la **importancia de cada variable** en la predicción. En la Figura 4.3 se observa que la generación renovable y fósil, junto con el PIB, son las variables más influyentes en la generación total.

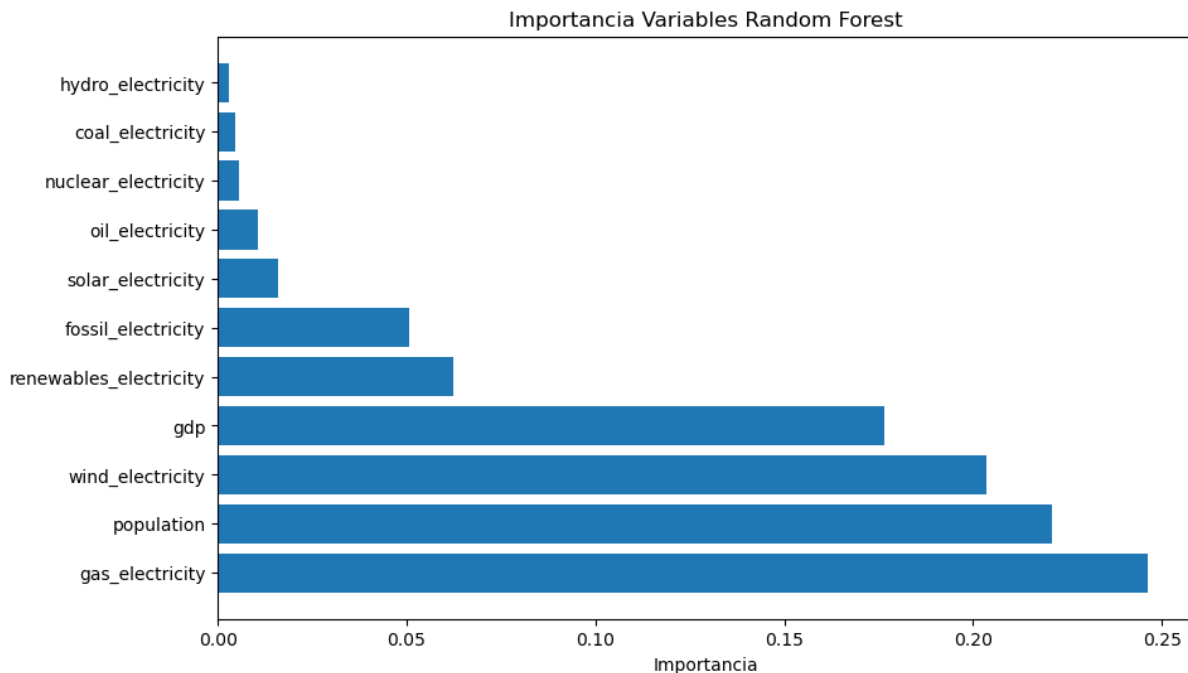


Figura 4.3: Importancia de variables según el modelo Random Forest.

### 4.3. Validación

Para evaluar el rendimiento de los modelos se utilizó **validación temporal** mediante `TimeSeriesSplit`, dividiendo el histórico en tres pliegues temporales y manteniendo el orden cronológico (evitando fugas de información del futuro hacia el pasado).

En cada pliegue se calcularon las métricas:

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

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

Los resultados obtenidos mostraron que:

- Prophet ofreció predicciones suaves y con buena captura de tendencia, pero algo menos preciso en años con variaciones abruptas.
- Random Forest mostró un ajuste más preciso en datos históricos y capacidad para capturar cambios no lineales, aunque sin producir intervalos de confianza.

Ambos modelos se entrenaron y guardaron en formato `.joblib` para su uso en el dashboard interactivo (`prophet_energy_model.joblib` y `rf_energy_model.joblib`).

En capítulos posteriores se compararán cuantitativamente sus resultados y se mostrará cómo se integran en la herramienta visual de análisis.

# Capítulo 5

## Resultados

En este capítulo se presentan los resultados obtenidos tras el entrenamiento y evaluación de los modelos desarrollados. Se analizan tanto las métricas cuantitativas como la interpretación de los patrones detectados, apoyándose en visualizaciones generadas en los *notebooks* y en el *dashboard* interactivo.

### 5.1. Desempeño de los Modelos

El rendimiento de los modelos se evaluó utilizando validación temporal y métricas de error medio absoluto (**MAE**) y raíz del error cuadrático medio (**RMSE**). Los resultados medios para los tres pliegues de validación se resumen en la Tabla 5.1.

Cuadro 5.1: Desempeño de los modelos en validación temporal

Modelo	MAE (GWh)	RMSE (GWh)
Prophet	7,93	9,31
Random Forest	4,87	7,99

A partir de estas métricas se observa que:

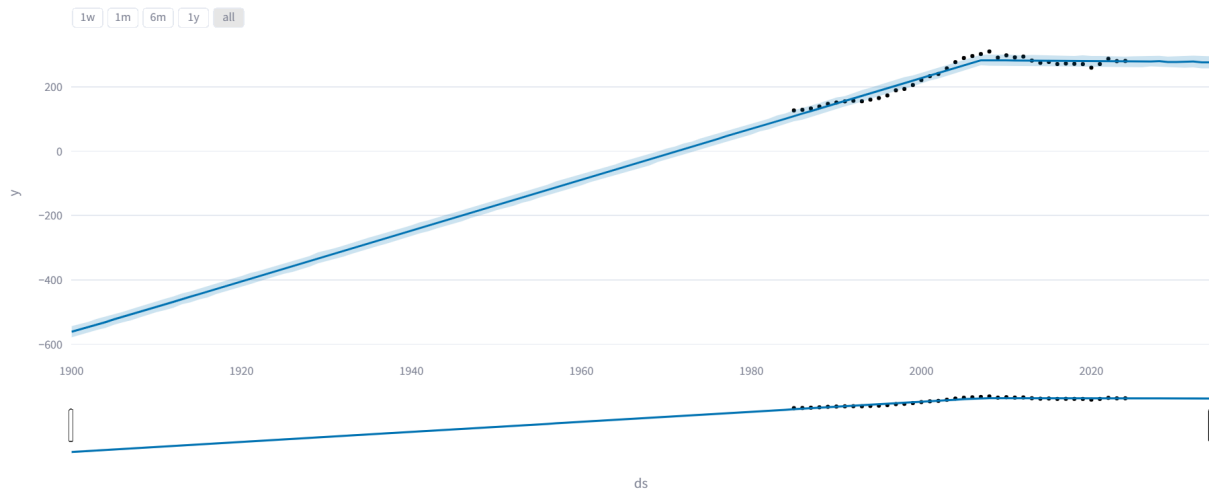
- El modelo **Random Forest** logra un error más bajo en promedio, especialmente en años con cambios bruscos.
- El modelo **Prophet** presenta un rendimiento ligeramente inferior en error, pero ofrece intervalos de confianza y un comportamiento más estable en predicciones a largo plazo.

En la Figura 5.1 se muestra un ejemplo de predicción a 10 años con Prophet, incluyendo el intervalo de confianza al 95 %. La Figura 5.2 compara los valores reales con las predicciones del modelo Random Forest en el conjunto de validación.



## Predicciones: Modelo seleccionado

Modelo **Prophet** para predicción de generación total eléctrica con intervalo de confianza.



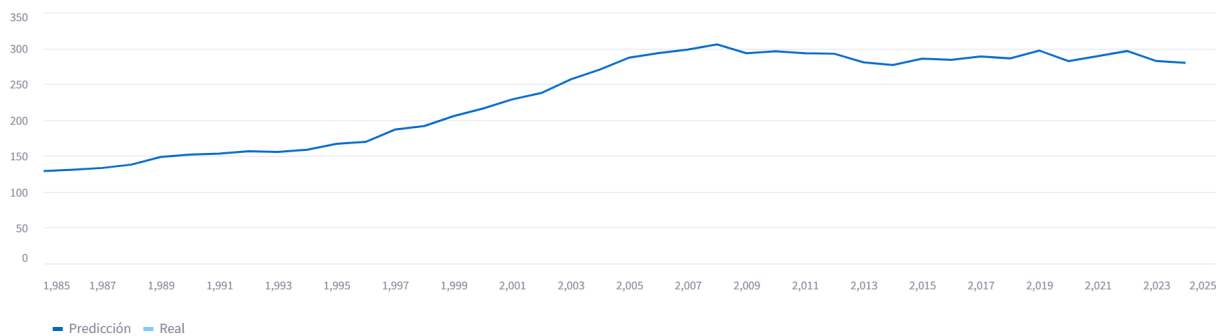
MAE (últimos 10 años): 7.93

RMSE (últimos 10 años): 9.31

Figura 5.1: Predicción de generación eléctrica total en España con Prophet.

## Predicciones: Modelo seleccionado

Modelo **Random Forest Regressor** para predicción multivariante de generación total eléctrica.



MAE: 4.87

RMSE: 7.99

Figura 5.2: Predicción de generación eléctrica total en España con Random Forest.

## 5.2. Interpretación de Resultados

El análisis de la importancia de variables en el modelo Random Forest (Figura 5.3) revela que:

- `gas_electricity` y `population` son los predictores más relevantes.
- El `gdp` contribuye significativamente, reflejando la relación entre crecimiento económico y consumo energético.

- Las fuentes específicas (**solar**, **wind**, **hydro**) tienen relevancia secundaria, pero ayudan a capturar variaciones anuales concretas.

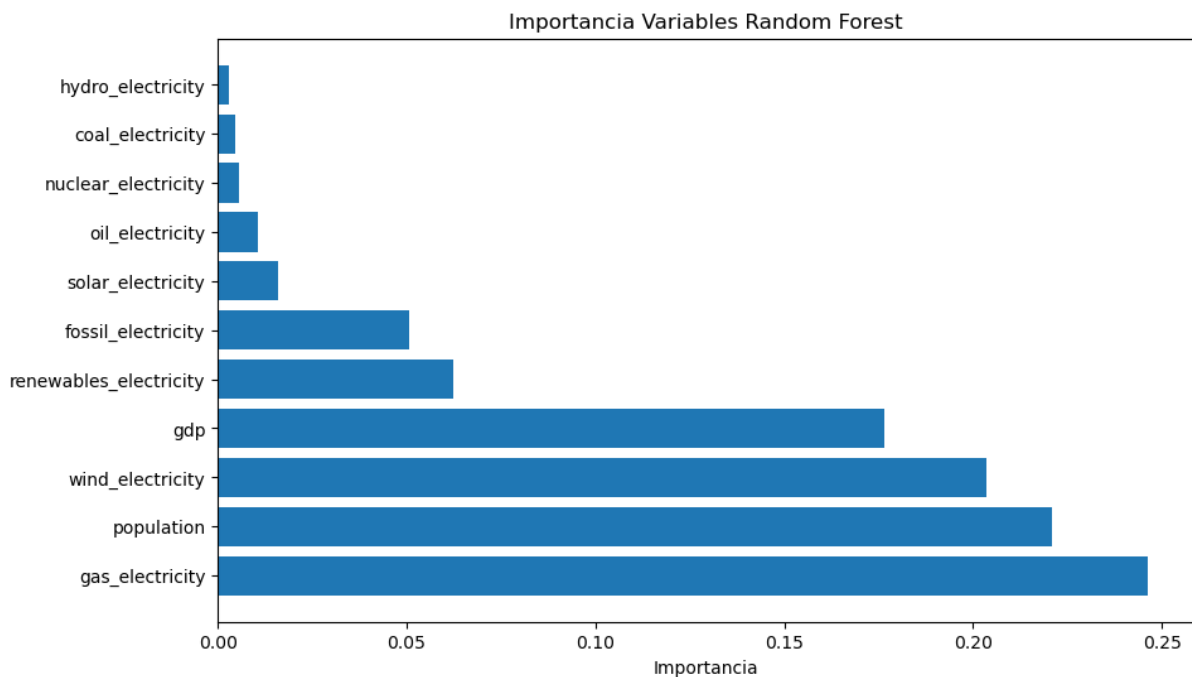


Figura 5.3: Importancia relativa de variables según el modelo Random Forest.

Estos resultados concuerdan con el análisis exploratorio previo (Capítulo 3), donde ya se observaban correlaciones altas entre la generación renovable/fósil y la generación total.

### 5.3. Limitaciones Observadas

Durante la evaluación se identificaron varias limitaciones:

- El dataset es anual, lo que limita la capacidad para capturar variaciones intraanuales o efectos estacionales cortos.
- Prophet no incorpora directamente variables exógenas en esta implementación, por lo que no aprovecha toda la información disponible.
- El modelo Random Forest, aunque preciso, no produce intervalos de confianza, dificultando la estimación de la incertidumbre en las predicciones.
- Los cambios abruptos por crisis energéticas o políticas (por ejemplo, cierres de centrales, variaciones drásticas en precios) son difíciles de anticipar para ambos modelos.

A pesar de estas limitaciones, los resultados obtenidos son robustos y muestran el potencial de los enfoques utilizados para anticipar tendencias y planificar estrategias energéticas a medio plazo.

# Capítulo 6

## Implementación

En este capítulo se describe cómo se llevó a cabo la implementación práctica del sistema de predicción de demanda energética, desde la integración de los modelos entrenados hasta su despliegue en un *dashboard* interactivo desarrollado con **Streamlit**.

### 6.1. Arquitectura del Proyecto

El proyecto se estructuró para separar claramente el flujo de trabajo en diferentes fases:

1. **Obtención y preprocesamiento de datos:** notebooks en la carpeta `notebooks/` para descarga, limpieza y análisis exploratorio.
2. **Entrenamiento de modelos:** notebooks dedicados a la construcción de Prophet y Random Forest.
3. **Persistencia de modelos:** almacenamiento en formato `.joblib` para carga rápida.
4. **Visualización y análisis interactivo:** aplicación en Streamlit `app.py`, ubicada en la carpeta `dashboard`.

### 6.2. Persistencia y Carga de Modelos

Tras el entrenamiento de Prophet y Random Forest, se guardaron los modelos en formato binario usando la librería `joblib`, lo que permite cargarlos posteriormente en el dashboard sin necesidad de reentrenar.

Listing 6.1: Guardado de modelos entrenados

```
import joblib

# Guardar Prophet
joblib.dump(model_prophet, '../src/prophet_energy_model.joblib')

# Guardar Random Forest
joblib.dump(model_rf, '../src/rf_energy_model.joblib')
```

En el dashboard, estos modelos se cargan al inicio para estar disponibles en memoria de forma inmediata:

Listing 6.2: Carga de modelos en el dashboard

```
@st.cache_resource(show_spinner=True)
def load_models():
    model_prophet = joblib.load('../src/prophet_energy_model.joblib')
    model_rf = joblib.load('../src/rf_energy_model.joblib')
    return model_prophet, model_rf

model_prophet, model_rf = load_models()
```

### 6.3. Carga y Preparación de Datos

El dashboard necesita el conjunto de datos limpio para poder mostrar tanto el histórico como las predicciones. Este se carga con **pandas** y se mantiene en memoria usando **@st.cache\_data** para mejorar el rendimiento.

Listing 6.3: Carga de datos limpios

```
@st.cache_data(show_spinner=True)
def load_data():
    df = pd.read_csv('../data/clean/energy_spain_clean.csv',
                     parse_dates=['date'])
    return df

df = load_data()
```

### 6.4. Interfaz en Streamlit

El *dashboard* se diseñó con una estructura de **barra lateral** para configurar opciones y un área principal para mostrar gráficos y métricas. El usuario puede seleccionar el modelo a emplear, activar o desactivar la visualización de correlaciones y mostrar la importancia de variables en Random Forest.

Listing 6.4: Opciones de la barra lateral

```
st.sidebar.header("Opciones de visualización y modelos")

model_option = st.sidebar.selectbox(
    "Selecciona modelo predictivo", ["Prophet", "Random Forest"]
)

show_corr = st.sidebar.checkbox("Mostrar mapa de correlacion",
                                value=True)
```

```
show_imp = st.sidebar.checkbox("Mostrar importancia variables RF",
                               value=True)
show_pred = st.sidebar.checkbox("Mostrar predicciones y comparativas", value=True)
```

## 6.5. Predicciones y Visualización

Para Prophet, se genera un DataFrame futuro con `make_future_dataframe`, se obtienen las predicciones y se muestran con el método `plot_plotly`, incluyendo intervalos de confianza.

Listing 6.5: Predicción con Prophet

```
if model_option == "Prophet":
    df_prophet = df.rename(columns={'date': 'ds', '
                                   electricity_generation': 'y'})[['ds', 'y']]
    future = model_prophet.make_future_dataframe(periods=10, freq='
        Y')
    forecast = model_prophet.predict(future)
    fig_forecast = plot_plotly(model_prophet, forecast)
    st.plotly_chart(fig_forecast, use_container_width=True)
```

En el caso de Random Forest, se realiza la predicción directamente sobre las variables explicativas y se muestran los valores reales frente a los predichos en un gráfico de líneas.

Listing 6.6: Predicción con Random Forest

```
if model_option == "Random Forest":
    features = ['renewables_electricity', 'fossil_electricity', '
               coal_electricity',
               'gas_electricity', 'nuclear_electricity', '
               oil_electricity',
               'hydro_electricity', 'solar_electricity', '
               wind_electricity',
               'gdp', 'population']
    df_rf = df.dropna(subset=features + ['electricity_generation'])
    .copy()
    X = df_rf[features]
    y_true = df_rf['electricity_generation']
    y_pred = model_rf.predict(X)
    comp_df = pd.DataFrame({'Real': y_true, 'Prediccion': y_pred},
                           index=df_rf['year'])
    st.line_chart(comp_df)
```

## 6.6. Distribución y Ejecución

La aplicación puede ejecutarse localmente con el comando:

Listing 6.7: Ejecución del dashboard

```
cd dashboard  
streamlit run app.py
```

El código y todos los recursos necesarios están organizados para que cualquier usuario pueda clonar el repositorio, instalar dependencias y ejecutar el dashboard sin modificar la estructura de carpetas.

## 6.7. Ventajas del Enfoque

Este tipo de implementación presenta varias ventajas:

- Permite separar el desarrollo del modelo del despliegue, facilitando la actualización de predicciones sin reentrenar desde cero.
- La interfaz es sencilla de usar incluso para usuarios sin conocimientos técnicos.
- El uso de Streamlit y joblib hace que el tiempo de carga sea mínimo, mejorando la experiencia de uso.

En el siguiente capítulo se expondrán las conclusiones del trabajo, junto con posibles líneas de mejora y extensión futura.

# Capítulo 7

## Conclusiones y Trabajo Futuro

En este capítulo se resumen los principales hallazgos del proyecto y se plantean posibles líneas de mejora que permitan ampliar sus capacidades y aplicabilidad.

### 7.1. Conclusiones

El presente trabajo ha desarrollado un sistema completo para la predicción de la demanda y generación energética en España, combinando análisis exploratorio de datos, modelado predictivo y visualización interactiva.

A partir de datos abiertos, se han implementado dos enfoques complementarios:

- **Prophet**, un modelo de series temporales que ofrece interpretabilidad y capacidad de proyección a largo plazo, con inclusión de intervalos de confianza.
- **Random Forest**, un modelo de *machine learning* multivariante que ha demostrado un mejor rendimiento en métricas como MAE y RMSE, aprovechando información de distintas fuentes energéticas y variables socioeconómicas.

Entre los logros más destacados se encuentran:

- Un *pipeline* reproducible que abarca desde la obtención de datos hasta el despliegue en un *dashboard* interactivo.
- Una arquitectura modular que permite sustituir o añadir modelos sin modificar la interfaz de usuario.
- Visualizaciones dinámicas que facilitan la interpretación de resultados y la exploración de tendencias históricas.

En términos de rendimiento, los modelos desarrollados muestran una capacidad predictiva robusta en un horizonte temporal de varios años. Esto resulta de gran utilidad para la planificación energética, especialmente en un contexto de transición hacia fuentes renovables y de creciente electrificación de la demanda.

Asimismo, el proyecto demuestra la viabilidad de integrar metodologías de análisis de datos e inteligencia artificial con herramientas de visualización web para crear aplicaciones útiles y accesibles a usuarios no técnicos.

## 7.2. Trabajo Futuro

Aunque los resultados son prometedores, existen múltiples áreas en las que el proyecto podría evolucionar:

1. **Mayor granularidad temporal:** trabajar con datos mensuales, diarios o incluso horarios permitiría capturar variaciones estacionales cortas y picos de demanda.
2. **Variables meteorológicas y contextuales:** integrar temperatura, radiación solar, velocidad del viento, indicadores de actividad económica y cambios regulatorios para mejorar la precisión.
3. **Nuevos modelos:** explorar arquitecturas basadas en redes neuronales recurrentes (LSTM, GRU) o modelos híbridos que combinen Prophet con predictores exógenos.
4. **Predicciones por tipo de fuente:** generar modelos independientes para solar, eólica, hidráulica, etc., y luego agregarlos para obtener la proyección total.
5. **Despliegue en la nube:** publicar el *dashboard* en plataformas como Streamlit Cloud, Heroku o AWS para acceso universal sin instalación local.
6. **Automatización del flujo de datos:** integrar un sistema de actualización automática que consulte nuevas observaciones y actualice las predicciones en tiempo real.

La combinación de estos desarrollos permitiría no sólo mejorar la calidad de las predicciones, sino también ampliar el alcance de la herramienta, convirtiéndola en un sistema de apoyo a la toma de decisiones para empresas, administraciones públicas y organismos de planificación energética.

En definitiva, este trabajo sienta una base sólida sobre la que se pueden construir soluciones más avanzadas, combinando datos abiertos, inteligencia artificial y visualización interactiva para contribuir a la optimización y sostenibilidad del sistema eléctrico.



# Bibliografía

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Hippert, H. S., Pedreira, C. E., and Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55.
- Hong, T., Pinson, P., and Fan, S. (2010). Short term electric load forecasting. *Encyclopedia of Operations Research and Management Science*, pages 1400–1408.
- Kowalska-Pyzalska, A. et al. (2023). Short-term electricity demand forecasting with machine learning: An extensive review. *Renewable and Sustainable Energy Reviews*, 184:113491.
- Marino, D. L., Amarasinghe, K., and Manic, M. (2016). Building energy load forecasting using deep neural networks. In *2016 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON)*, pages 7046–7051. IEEE.
- Ritchie, H., Roser, M., and Rosado, P. (2023). Our world in data: Energy dataset. <https://github.com/owid/energy-data>. Accedido: Julio 2025.
- Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1):37–45.
- Wang, X. et al. (2019). Review on deep learning techniques for the forecasting of energy demand in smart grids. *Applied Energy*, 253:113637.
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081.