



Universidad Nacional
de La Matanza



Especialización en Ciencia de Datos

*Predicción de Demanda Eléctrica en Argentina mediante
Técnicas de Ciencia de Datos*

*Trabajo Final Integrador presentado para la obtención del título de
especialista en Ciencia de Datos*

*Autor: Emanuel Damian Medina
(Ingeniero Industrial e Ingeniero en Informática)*

*Tutor: Blanca Patricia Ynsfran
(Magister en Gestión Ambiental)*

Fecha de entrega: septiembre de 2025

Contenido

Agradecimientos	4
Capítulo 1. Resumen	4
Capítulo 2. Marco conceptual	5
Capítulo 3. Antecedentes	5
Capítulo 4. Planteo del problema.....	6
4.1 Problema	6
4.2 Justificación del estudio	6
4.3 Alcance.....	7
4.4 Limitaciones.....	7
4.5 Hipótesis y variables	7
4.5.1 Variable dependiente:	8
4.5.2 Variables independientes:	8
4.5.3 Variables de control:	8
Capítulo 5. Objetivos	8
5.1 Objetivo general.....	8
5.1 Objetivos específicos	8
Capítulo 6. Metodologías técnicas.....	9
6.1 Introducción	9
6.2 Reproducibilidad y escalabilidad del modelo	9
6.3 Herramientas	10
Capítulo 7. Desarrollo metodológico.....	10
7.1 Construcción del Dataset y preparación de datos	11
7.1.1 Transformación de la demanda eléctrica	12
7.1.2 Procesamiento de datos de importación y exportación.....	12
7.1.3 Obtención de variables de contexto: feriados y clima	12
7.1.4 Pipeline de integración y Dataset final	12
7.2 Resultados preliminares del Dataset	13
7.2.1 Demanda eléctrica por provincia	13
7.2.2 Importación y exportación de energía.....	13
7.2.3 Dataset final unificado	13
7.3 Análisis exploratorio de la demanda eléctrica	13

7.3.1	Series temporales horarias	14
7.3.2	Implicancias del horario en la demanda eléctrica	14
7.3.3	Distribución de la demanda.....	15
7.3.4	Conclusión del análisis exploratorio	19
7.4	Preparación de features y preprocesamiento.....	19
7.4.1	Preparación de features	20
7.4.2	Preprocesamiento: Codificación de variables categóricas para Pipeline	20
7.5	Entrenamiento de modelos.....	20
7.5.1	Random Forest Regressor	20
7.5.2	Gradient Boosting Regressor	21
7.5.3	Red Neuronal IRBDNN con bloques residuales.....	21
7.5.4	Red Neuronal IRBDNN Pura (sin bloques residuales).....	22
7.6	Evaluación de modelos	22
7.7	Selección del modelo final.....	25
7.7.1	Rendimiento cuantitativo	25
7.7.2	Análisis visual	25
7.7.3	Robustez y generalización	25
7.7.4	Comparación con otros modelos.....	26
7.7.5	Justificación de la elección	26
7.8	Interpretación del Modelo: Importancia de las Variables.....	27
Capítulo 8. Futuros estudios.....		27
Capítulo 9. Conclusiones y recomendaciones.....		28
9.1	Conclusión	28
9.2	Recomendaciones prácticas	28
Capítulo 10. Apéndice técnico.....		29
10.1	Código fuente.....	29
10.1.1	Preparación de los Dataset y ejecución.....	29
10.2	Fuentes de Datos y Datasets Utilizados	30
Referencias bibliográficas.....		31

Agradecimientos

Quiero expresar mi sincero agradecimiento a la Universidad Nacional de La Matanza por brindarme la oportunidad de participar en esta Especialización en Ciencia de Datos, una formación de excelencia que me permitió integrar conocimientos teóricos y prácticos en un proyecto aplicado de alto impacto.

Asimismo, agradezco especialmente al tutor asignado, por su tiempo, acompañamiento y valiosos aportes durante el desarrollo de este trabajo final.

Capítulo 1. Resumen

La planificación de la demanda eléctrica es una de las tareas más críticas para un funcionamiento eficiente y sostenible de la matriz¹ energética nacional. Factores climáticos y estructurales inciden significativamente sobre el consumo eléctrico, disponer de modelos predictivos que integren múltiples fuentes de información resulta indispensable para poder tener herramientas de soporte para todo lo relacionado a la gestión de la energía eléctrica.

Este trabajo propone un enfoque basado en técnicas de ciencia de datos para predecir la demanda eléctrica horaria a nivel nacional, incorporando variables climáticas, energéticas (saldo de intercambio energético con países limítrofes)² y variables temporales. Se emplearán distintos modelos de Deep Learning que serán evaluados con distintas métricas. Los resultados obtenidos buscarán demostrar cual es el mejor método para anticipar los picos de consumo en determinados horarios, con determinadas condiciones climáticas y/o tipo de día, lo cual va a contribuir en una mejora en la toma de decisiones al momento de la planificación operativa y estratégica tanto del sistema eléctrico SADI³, como así también de las redes de transmisión⁴ y distribución⁵ de las distintas regiones del país.

¹ La matriz energética se refiere a la combinación de fuentes de energía utilizadas para satisfacer la demanda energética. En Argentina, la matriz energética se caracteriza por una alta dependencia de los hidrocarburos.

² El saldo de intercambio energético con países limítrofes es la diferencia entre la cantidad de energía que un país exporta y la cantidad que importa. Un saldo positivo significa que el país exporta más energía de la que importa, mientras que un saldo negativo indica que importa más de lo que exporta.

³ Sistema Argentino de Interconexión Eléctrica. Sistema de red interconectado, de 500 kw, que atraviesa el país desde las provincias del norte hasta Santa Cruz. La única provincia que no está unida a ese sistema es Tierra del Fuego.

⁴ La transmisión de energía eléctrica es el proceso de transporte de electricidad desde las centrales generadoras hasta las subestaciones eléctricas por medio de líneas de alta tensión, donde se transforma y se distribuye a los consumidores finales.

⁵ La distribución de energía eléctrica es la fase final donde la energía se lleva desde las subestaciones hasta los usuarios finales. Este proceso implica la transformación y adaptación de la tensión de la energía eléctrica para su uso seguro y eficiente en los diferentes puntos de consumo.

Capítulo 2. Marco conceptual

Es esencial planificar la generación y distribución de manera eficiente, optimizando los costos de operación y mantenimiento de las redes, como así también asegurar un suministro eléctrico acorde a las necesidades. Como afirma Nti et al. (2020), *“La previsión de carga a corto plazo ha atraído una atención sustancial en los sistemas de energía debido a su importancia en la gestión de la energía en tiempo real”*.

Los enfoques de predicción han evolucionado desde modelos estadísticos tradicionales hacia técnicas más sofisticadas basadas en Machine Learning, Deep Learning y procesamiento de series temporales. La disponibilidad de datos abiertos relacionados a la generación y consumo, como así también de fuentes de variables exógenas han potenciado la necesidad de desarrollos de modelos predictivos más precisos y contextualizados.

Como señalan Hammad et al. (2020), *“Una previsión precisa requiere una selección adecuada de características, elección de modelos y técnicas de evaluación de errores”*, por lo que la Ciencia de Datos se posiciona como una herramienta clave e indispensable para optimizar la gestión energética actual. De esta manera, como consecuencia se garantiza entre otras cosas el cumplimiento satisfactorio de los indicadores internacionales de SAIFI⁶ y SAIDI⁷.

Capítulo 3. Antecedentes

Diversos estudios internacionales han abordado el problema de la predicción de consumo de energía eléctrica. Como afirman Nti et al. (2020), *“La previsión de la carga eléctrica desempeña un papel fundamental en la planificación y el funcionamiento eficientes de la energía”*, lo que resalta la necesidad de modelos robustos para evitar pérdidas por fallas en redes eléctricas y sobrecargas innecesarias. Dichos autores presentan una revisión sistemática que clasifica enfoques de predicción según el horizonte temporal y las fuentes de datos utilizadas.

Desde una perspectiva metodológica, Rostum et al. (2020), *“La previsión precisa de la carga eléctrica es vital para el funcionamiento seguro y estable del sistema eléctrico”*. Ofrecen una visión panorámica sobre las principales técnicas empleadas históricamente en la predicción de demanda eléctrica, incluyendo ventajas y limitaciones.

⁶ (System Average Interruption Frequency Index). Mide la frecuencia con la que los clientes experimentan interrupciones del suministro eléctrico en un período determinado. Se expresa en la cantidad de interrupciones por cliente al año.

⁷ (System Average Interruption Duration Index). Mide la duración total de las interrupciones del servicio eléctrico para el cliente promedio en un período determinado. Se expresa en unidades de tiempo, como horas por cliente al año.

Finalmente, Solyali (2020) concluye que *“Las técnicas de aprendizaje automático, como el árbol de decisión y el aumento de gradiente, son eficaces para la previsión de carga tanto a corto como a largo plazo,”* validando su inclusión en este trabajo. Se realizó un análisis comparativo de modelos de Machine Learning aplicados a series temporales de demanda eléctrica en Chipre, validando la efectividad de modelos basados en árboles y redes profundas.

En este marco, CAMMESA⁸ provee datos de demanda eléctrica horaria y de saldo de intercambio energético con países limítrofes, mientras que Open-Meteo⁹ ofrece información climática gratuita según las coordenadas geográficas de la región en análisis. Nager.date¹⁰ proporciona información sobre los días feriados en Argentina. Esta combinación habilita estudios empíricos basados en datos públicos, replicables y relevantes.

Capítulo 4. Planteo del problema

4.1 Problema

La demanda eléctrica en Argentina se ve significativamente influenciada por factores exógenos. Sin un análisis riguroso de estas variables, tanto la planificación operativa como estratégica del sistema energético enfrenta riesgos de desbalance, aumento de costos operativos, de mantenimiento y vulnerabilidad ante eventos por consumo.

En este contexto nos podemos plantear el siguiente interrogante, *¿Cómo influyen las variables climáticas y el saldo de intercambio energético internacional sobre la demanda eléctrica horaria?*

4.2 Justificación del estudio

La creciente demanda energética, los fenómenos climáticos extremos y la variabilidad en los intercambios internacionales de energía eléctrica exige el desarrollo e implementación de herramientas aún más eficientes y precisas que las actuales para la predicción de consumo eléctrico. Como explica Rostum et al. (2020), *“La previsión de carga es esencial para mantener una red eléctrica equilibrada y estable, especialmente en presencia de condiciones variables de demanda y oferta”*

⁸ CAMMESA (Compañía Administradora del Mercado Mayorista Eléctrico Sociedad Anónima) es una compañía argentina encargada de operar el mercado eléctrico mayorista de Argentina.

⁹ Open-Meteo es una API meteorológica de código abierto y ofrece acceso gratuito para uso no comercial.

¹⁰ Nager.date es un servicio API REST gratuito que proporciona información sobre días festivos públicos y días laborables en más de 100 países. Permite a los usuarios acceder a datos de fechas, países y regiones específicas para planificar horarios y eventos.

Eventos de olas de calor y/o frío provocan picos de consumo que superan la capacidad instalada de los sistemas de Transmisión y Distribución, generando cortes y fallas. Una planificación ineficiente también compromete la calidad del servicio suministrado de energía eléctrica.

El presente estudio propone el desarrollo de distintos modelos predictivos basados en Ciencia de Datos que permitan anticipar como se comportará la demanda eléctrica horaria a partir de variables exógenas. La elección de fuentes de información abiertas y accesibles como CAMMESA, Nager.date y Open-Meteo no solo facilita la replicabilidad del trabajo, sino que también promueve el uso de datos públicos y confiables para mejorar la toma de decisiones en políticas energéticas nacionales.

4.3 Alcance

Este trabajo se enfoca en la predicción de la demanda eléctrica horaria a nivel nacional en Argentina, considerando datos del Sistema Argentino de Interconexión (SADI) durante el período 2022–2024. El análisis integra variables climáticas, energéticas y temporales mediante técnicas de aprendizaje supervisado.

4.4 Limitaciones

En el presente documento no se tienen en cuenta variables económicas, demográficas o tarifarias para una simplificación del modelo. La precisión del modelo obtenido está sujeta a la calidad, cantidad y continuidad de las fuentes de datos abiertas propuestas.

No se incorporan técnicas específicas de análisis de series temporales, como ser una descomposición estacional. No se incorporan pruebas estadísticas para determinar si las propiedades estadísticas de una serie temporal permanecen constantes en el tiempo., lo cual representa una limitación significativa para un problema inherentemente temporal.

4.5 Hipótesis y variables

Como hipótesis para el desarrollo del presente trabajo se plantea que la temperatura promedio horaria y el saldo de intercambio energético internacional inciden significativamente en la variabilidad de la demanda eléctrica horaria en Argentina.

4.5.1 Variable dependiente:

- Demanda eléctrica horaria (MW¹¹): variable de tipo cuantitativa continua que indica el consumo de energía eléctrica, medido en megavatios, en intervalos de una hora.

4.5.2 Variables independientes:

- Temperatura promedio horaria (°C): variable de tipo cuantitativa continua que indica el valor de la temperatura ambiente registrada, que se interpreta que influye sobre el uso de elementos de climatización (calefacción o refrigeración). Medida en grados Celsius.
- Saldo comercial energético (MWh¹²): variable de tipo cuantitativa continua que indica la diferencia entre la energía eléctrica importada y exportada por Argentina. Puede compensar faltantes o excedentes en la demanda. Medido en megavatios por hora.

4.5.3 Variables de control:

- Hora del día: variable de tipo cualitativa discreta que identifica la franja horaria en la que se produce el consumo eléctrico. Incide en los picos de demanda.
- Día del mes: variable de tipo cualitativa ordinal que identifica el día de un mes corresponde.
- Tipo de día: variable de tipo cualitativa nominal que indica si es un día hábil o no.
- Feriado: variable de tipo cualitativa nominal que identifica si el día es feriado nacional o no, lo cual se estima que modifica significativamente la demanda.

Capítulo 5. Objetivos

5.1 Objetivo general

Desarrollar mediante el uso de técnicas de ciencia de datos un modelo de predicción de la demanda eléctrica horaria en Argentina que integre datos abiertos de consumo, clima e intercambios energéticos internacionales.

5.1 Objetivos específicos

- Recolectar y unificar datos históricos de demanda eléctrica, clima e intercambios energéticos.
- Explorar y visualizar patrones de comportamiento de la demanda.

¹¹ Un megavatio (MW) es una unidad de medida de potencia equivalente a un millón de vatios (1.000.000 W) o 1000 kilovatios (kW). Se utiliza para medir la capacidad de generación de electricidad a gran escala, como la de centrales eléctricas.

¹² Un megavatio-hora (MWh) es una unidad de medida de energía que representa la cantidad de energía consumida o generada por un megavatio (MW) de potencia durante una hora. En otras palabras, es la energía producida o consumida cuando se utiliza un megavatio de potencia durante 60 minutos.

- Generar variables adicionales para mejorar la capacidad predictiva del modelo.
- Entrenar y evaluar el desempeño de distintos modelos de Deep Learning con métricas estadísticas como MAE, RMSE, R^2 y MSE; entre los modelos a implementar se destacan Gradient Boosting, Random Forest, IRBDNN y IRBDNN (residual).
- Documentar el proceso en un Pipeline reproducible que pueda ser reutilizado en otros contextos o escalado a nivel regional.

Capítulo 6. Metodologías técnicas

6.1 Introducción

El trabajo se desarrolla bajo un enfoque de análisis cuantitativo, basado en técnicas de Ciencia de Datos aplicadas a datos temporales. Se utilizará una metodología estructurada en cinco etapas principales:

- Extracción y unificación de datos: recopilación de archivos históricos desde CAMMESA y APIs públicas como Open-Meteo y Nager.date.
- Preprocesamiento: limpieza, normalización, creación de nuevas variables (como rezagos y rolling), imputación de valores faltantes.
- Análisis exploratorio: visualización de tendencias, correlaciones y estacionalidades mediante herramientas gráficas.
- Modelado predictivo: entrenamiento de modelos regresivos (Gradient Boosting, Random Forest, IRBDNN y IRBDNN residual) sobre datos divididos en entrenamiento/test.
- Evaluación y selección: comparación del rendimiento mediante métricas como MAE, RMSE, MSE y R^2 para elegir el modelo más robusto y confiable.

6.2 Reproducibilidad y escalabilidad del modelo

Uno de los objetivos fundamentales en el desarrollo de soluciones basadas en Ciencia de Datos es garantizar la reproducibilidad del análisis, es decir, que cualquier investigador o profesional pueda replicar los resultados obtenidos a partir del mismo conjunto de datos y código. Para ello, el presente trabajo se estructurará como un Pipeline automatizado y documentado.

El Pipeline se compone de etapas secuenciales que incluyen:

- Carga de datos: desde archivos CSV estandarizados con formatos conocidos.

- Integración y limpieza: combinación de fuentes heterogéneas (clima, demanda, feriados, intercambios), imputación de valores faltantes y estandarización de variables temporales. Integrándolos en un nuevo Dataset con las variables necesarias para el análisis.
- Ingeniería de características: creación de variables derivadas necesarias con el propósito de optimizar el análisis de los datos.
- Entrenamiento de modelos: ejecución automática de cuatro algoritmos (Gradient Boosting, Random Forest, IRBDNN y IRBDNN residual), con partición de datos y evaluación interna.
- Evaluación y visualización: cálculo de métricas de error (MAE, RMSE, MSE y R^2) y representación gráfica de las predicciones frente a los valores reales.

La estructura modular del Pipeline permite escalar el análisis a otras regiones o países simplemente cambiando los archivos de entrada, sin necesidad de modificar el código interno. Esta organización promueve tanto la reutilización como la extensión del modelo, lo que es clave en contextos de implementación práctica o futuros desarrollos institucionales. Asimismo, el uso de herramientas de código abierto asegura la accesibilidad y portabilidad del trabajo.

6.3 Herramientas

A continuación, se detallan las herramientas utilizadas:

- Lenguaje de programación: Python 3.10
- Entorno de desarrollo: Jupyter Notebook / Google Colab
- Gestión de versiones y reproducibilidad: Archivos en formato .csv organizados por año y script unificado para la automatización del proceso.

Capítulo 7. Desarrollo metodológico

La metodología sistemática implementada se encuentra estructurada en cinco fases secuenciales y complementarias.

- Preparación de datos, donde se ejecuta la construcción del Dataset mediante la transformación de datos de demanda eléctrica, de importación/exportación, y la integración de variables contextuales como feriados y condiciones.
- Análisis exploratorio exhaustivo, incluyendo el estudio de series temporales horarias, identificación de patrones de consumo y análisis de distribuciones estadísticas.

- Preparación de features mediante técnicas de feature engineering y preprocesamiento, incluyendo la codificación de variables.
- Modelización implementando cuatro algoritmos distintos.
- Evaluación cuantitativa y cualitativa de los modelos, incorporando análisis de performance, robustez y capacidad de generalización para la selección del modelo óptimo, culminando con la interpretación de la importancia de variables predictoras.

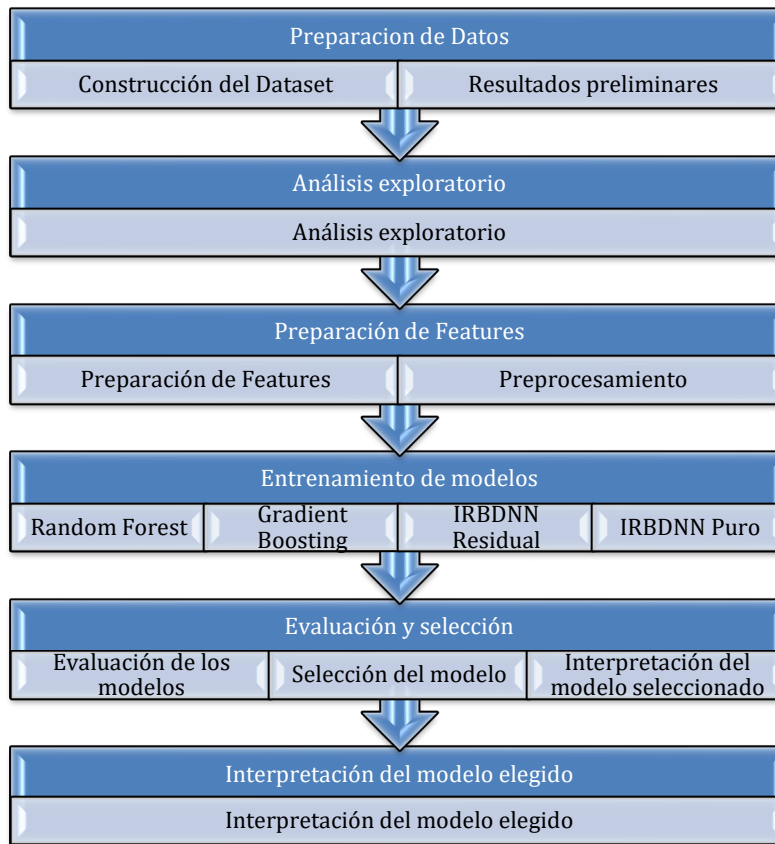


Gráfico 1 - Diagrama de flujo metodológico para el desarrollo de modelos de predicción de demanda eléctrica.

7.1 Construcción del Dataset y preparación de datos

Para llevar a cabo la predicción de la demanda eléctrica horaria en Argentina, se diseñó un proceso de integración y preparación de datos a partir de fuentes públicas. La recolección de datos se realizó de distintas fuentes confiables, utilizando Python como entorno de desarrollo y Google Colab para su ejecución reproducible.

7.1.1 Transformación de la demanda eléctrica

Se convierte la base de datos de formato ancho a formato largo, filtrando únicamente los campos de interés y agregando variables contextuales como *TipoDia* (hábil, no hábil o feriado) y *DiaSemana* (lunes a domingo). Además, se aplica un redondeo uniforme sobre los valores de demanda para normalizar magnitudes y facilitar la interpretación analítica. El Dataset resultante incluye las columnas *TipoDia*, *DiaSemana*, *Fecha*, *Hora*, *Provincia*, *Demanda*.

7.1.2 Procesamiento de datos de importación y exportación

Para complementar el análisis energético, se incorporaron los datos de importación y exportación de energía a nivel horario. Esta información se extrajo también desde la plataforma de CAMMESA, filtrando exclusivamente el período desde el 1 de enero de 2022 hasta el 31 de diciembre de 2024. El Dataset resultante incluye las columnas *Fecha*, *Hora*, *Importacion*, *Exportacion*.

7.1.3 Obtención de variables de contexto: feriados y clima

Feriatos: se generó un calendario oficial de feriados nacionales y regionales mediante la API Nager.Date. Permitiendo automatizar la lista de feriados para cada año del período analizado.

Datos climáticos: se obtuvieron desde la API Open-Meteo, generando series horarias para cada provincia con variables meteorológicas clave, como ser *TempC* (temperatura en grados Celsius), *Humedad*, y *Precipitación*. El resultado de esta etapa son archivos tabulares homogéneos con las columnas *Fecha*, *Hora*, *Provincia*, *TempC*, *Humedad*, *Precipitacion*.

7.1.4 Pipeline de integración y Dataset final

El proceso incluye:

- Conversión de fechas y horas a formato *datetime* para garantizar coherencia temporal.
- Generación de claves únicas (*Fecha_Hora*, *clave_provincia_hora*) para unir series horarias de demanda y clima por provincia.
- Integración de datos de importación/exportación replicando saldos horarios para cada provincia.
- Creación de la columna booleana *EsFeriado*, que marca “SI” o “NO” en función de la coincidencia con el calendario de feriados.
- Validación de duplicados, imputación de valores nulos y reordenamiento de columnas finales.

El Dataset final contiene las variables estructuradas *Fecha*, *Hora*, *Provincia*, *TipoDia*, *DiaSemana*, *EsFeriado*, *Demanda*, *TempC*, *Humedad*, *Precipitacion*, *Importacion*, *Exportacion*.

Este archivo se exporta en formato CSV y se respalda en Google Drive, garantizando su disponibilidad para entrenamiento de modelos, evaluación y comparación de resultados.

7.2 Resultados preliminares del Dataset

7.2.1 Demanda eléctrica por provincia

La variable dependiente Demanda fue redondeada al múltiplo de 100 más cercano para homogeneizar valores, obteniéndose las siguientes estadísticas descriptivas de la variable demanda.

- Promedio: 1439,97 MWh
- Máximo: 27.500 MWh

Se observa desde esta etapa preliminar que la Provincia de Buenos Aires concentra más del 60% de la demanda total, motivo por el cual se analizará de forma diferenciada en apartados posteriores.

7.2.2 Importación y exportación de energía

El Dataset resultante en esta etapa contiene 26.304 registros, obteniéndose las siguientes estadísticas descriptivas.

- Importación promedio: 654,06 MWh (máx.: 3.328,29 MWh).
- Exportación promedio: 41,78 MWh (máx.: 3.427,73 MWh).

Se constata que las importaciones suelen ser mayores y más frecuentes que las exportaciones, lo que refleja la necesidad estructural del SADI de reforzar la cobertura en horarios críticos.

7.2.3 Dataset final unificado

La magnitud y granularidad de los datos permiten capturar patrones horarios, estacionales y geográficos de la demanda eléctrica, integrando además factores exógenos como clima, feriados y flujo energético internacional.

7.3 Análisis exploratorio de la demanda eléctrica

Para comprender la dinámica de la demanda eléctrica, se desarrollaron funciones de visualización específicas en Python que permiten analizar la variabilidad horaria, estacional y por segmentos de contexto.

7.3.1 Series temporales horarias

En el siguiente gráfico se puede observar picos de demanda coincidentes con períodos de altas temperaturas, reflejando la sensibilidad climática del consumo eléctrico. Además, se evidencia la periodicidad horaria, con rangos de variabilidad mayores durante horarios pico de actividad productiva y residencial. Para los años 2023 y 2024 se observa un comportamiento similar

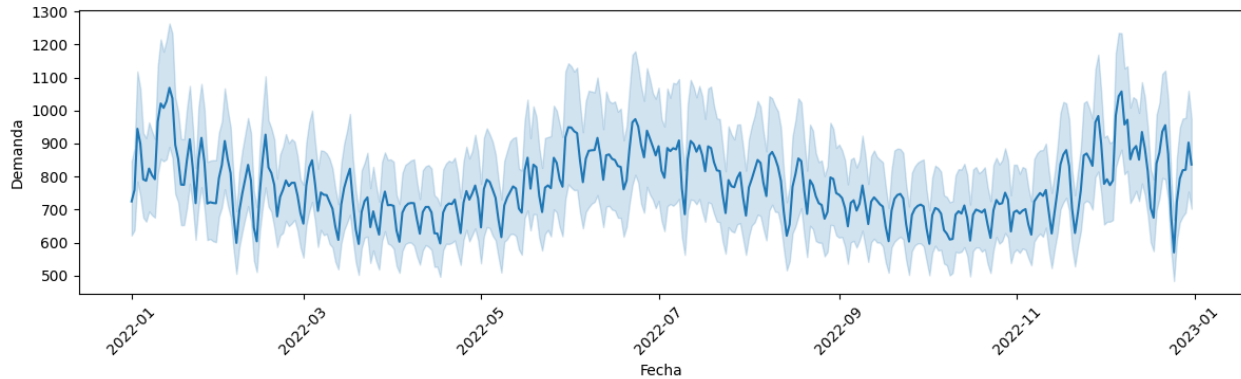


Gráfico 2 – Evolución de la demanda eléctrica horaria del año 2022. La banda sombreada refleja la dispersión de los valores en torno al promedio horario, evidenciando la volatilidad en ciertos períodos.

7.3.2 Implicancias del horario en la demanda eléctrica

La demanda eléctrica horaria constituye una de las variables más críticas para la gestión del sistema eléctrico argentino.

Identificando los picos horarios permite anticipar la necesidad de despacho de generación adicional, lo que evita riesgos de sobrecargas. En horarios de baja demanda, es posible programar los mantenimientos preventivos, reduciendo el riesgo de fallas. Permite realizar una asignación eficiente de recursos y repuestos para realizar las tareas de mantenimiento preventivo y predictivo.

Predecir la evolución horaria de la demanda, permite programar intervenciones sobre equipos críticos con mayor certeza, optimizando la disponibilidad de repuestos y cuadrillas técnicas. Esto repercute directamente en la reducción de costos operativos y en la continuidad del servicio.

Con relación al balance energético e importación/exportación, el horario influye en la magnitud de los intercambios energéticos internacionales. En particular, la cobertura de picos de demanda mediante importaciones es una práctica frecuente, y anticipar estos eventos ayuda a negociar contratos y optimizar el uso del interconectado regional.

Como se puede observar en el siguiente gráfico, se revela una relación compleja y no lineal entre la temperatura ambiente y la demanda de energía eléctrica. En temperaturas bajas, se observa un incremento progresivo en el consumo energético, lo cual refleja la necesidad de calefacción durante los períodos fríos. Por otro lado, en temperaturas elevadas, se presenta un segundo pico de demanda que alcanza niveles similares, asociado principalmente al uso intensivo de sistemas de refrigeración y aire acondicionado. Esta relación bimodal es fundamental para la planificación energética y la gestión de la demanda estacional.

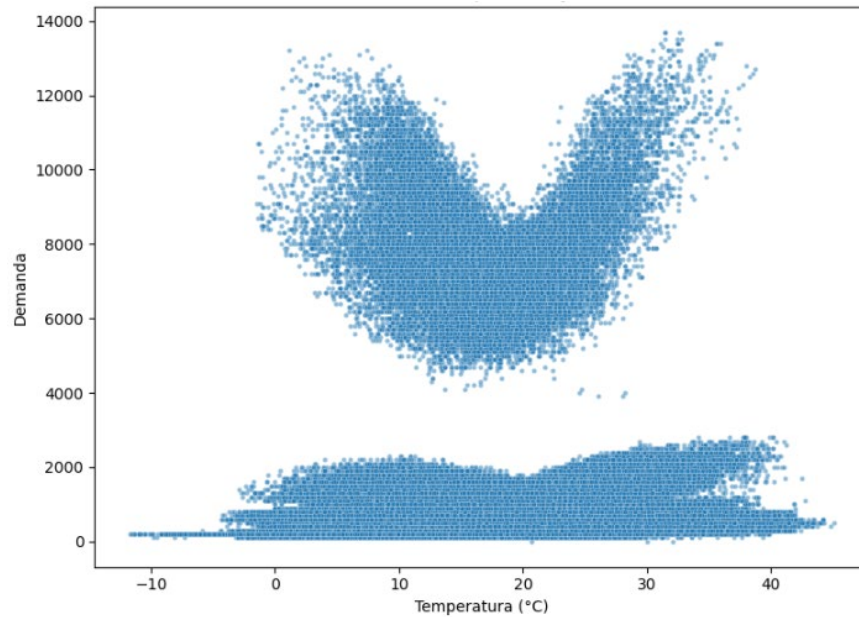


Gráfico 3 - Relación entre temperatura y demanda.

7.3.3 Distribución de la demanda

Para un análisis preliminar se excluyó Buenos Aires, ya que presenta valores de consumo notablemente superiores al resto de las provincias, lo que podría desvirtuar la visualización conjunta y ocultar patrones relevantes en regiones de menor demanda.

El gráfico de densidad muestra la distribución probabilística de la demanda energética de las principales provincias del sistema interconectado argentino. La representación evidencia marcadas diferencias regionales en los patrones de consumo eléctrico. Buenos Aires presenta una distribución bimodal distintiva, con concentraciones significativas tanto en rangos de baja demanda como en rangos de alta demanda, reflejando la diversidad de escenarios de consumo en la región más poblada del país.

Córdoba muestra una distribución más concentrada con picos de consumo, indicando un patrón de consumo más estable y predecible. Mendoza presenta características similares a Córdoba, pero con menor amplitud, sugiriendo niveles de demanda relativamente uniformes concentrados en rangos bajos a moderados. Santa Fe exhibe una distribución intermedia con cierta dispersión, posicionándose entre los patrones de Buenos Aires y las provincias del interior.

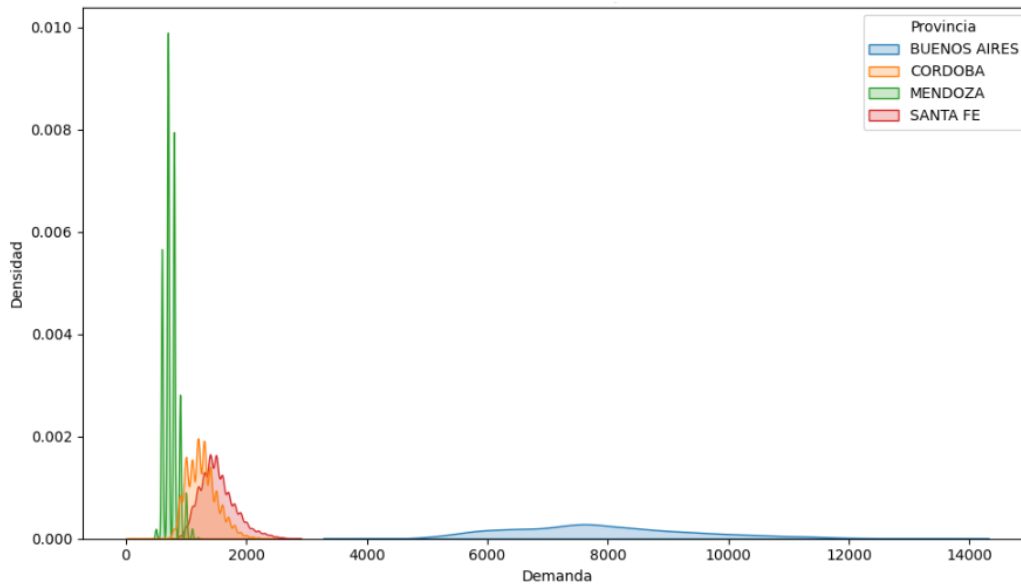


Gráfico 4 - Distribución de densidad de la demanda por provincias seleccionadas, mostrando la heterogeneidad regional en los patrones de consumo eléctrico.

La notable diferencia en las escalas de densidad refleja no solo las diferencias poblacionales sino también los distintos perfiles económicos e industriales de cada región. Esta heterogeneidad regional es fundamental para el diseño de estrategias diferenciadas de gestión de la demanda y planificación de la infraestructura de transmisión, permitiendo optimizar la operación del sistema eléctrico considerando las particularidades de cada jurisdicción. Tema que no será tratado en este documento, pero si se menciona en el final del mismo recomendaciones y conclusiones preliminares para futuros estudios.

Analizando la curva de distribución, se puede observar una cola extendida hacia valores superiores, con frecuencias decrecientes exponencialmente a medida que aumenta la demanda. Esta característica indica que los picos de alta demanda constituyen eventos relativamente infrecuentes pero críticos para la planificación del sistema. La presencia de valores en el extremo superior representa los períodos de máximo estrés del sistema, típicamente asociados a condiciones climáticas extremas o eventos de alta demanda sincronizada.

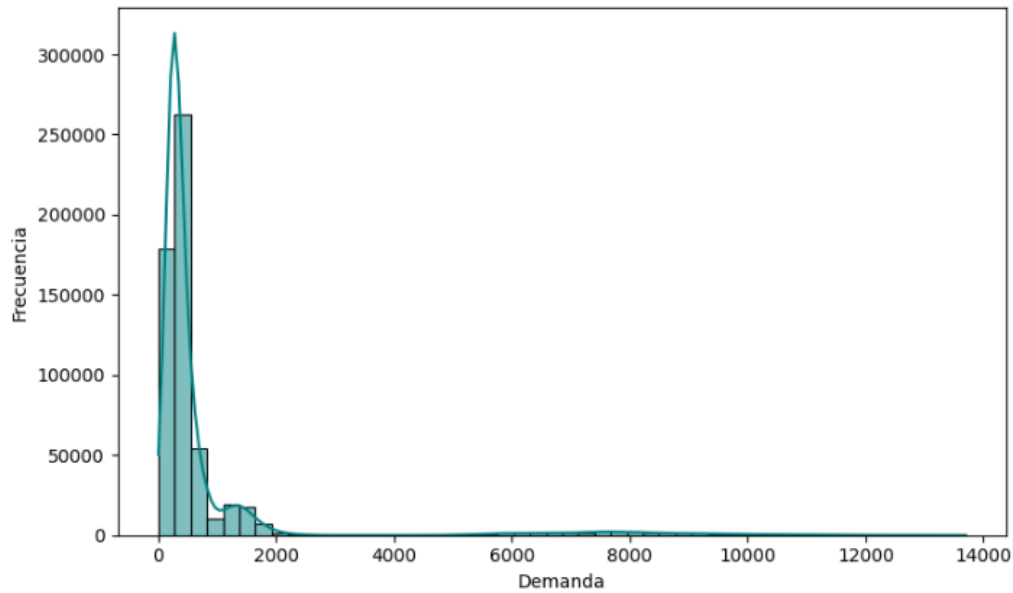


Gráfico 5 - Distribución de la demanda eléctrica. Histograma de distribución de frecuencias de la demanda eléctrica total del sistema, evidenciando la concentración estadística del consumo energético y la identificación de valores atípicos.

Esta distribución asimétrica es característica de los sistemas eléctricos, donde la demanda base constituye la mayor proporción del tiempo operativo, mientras que los picos de demanda requieren capacidad de generación adicional que permanece en reserva la mayor parte del tiempo. Esta información es fundamental para el dimensionamiento económico de la capacidad instalada.

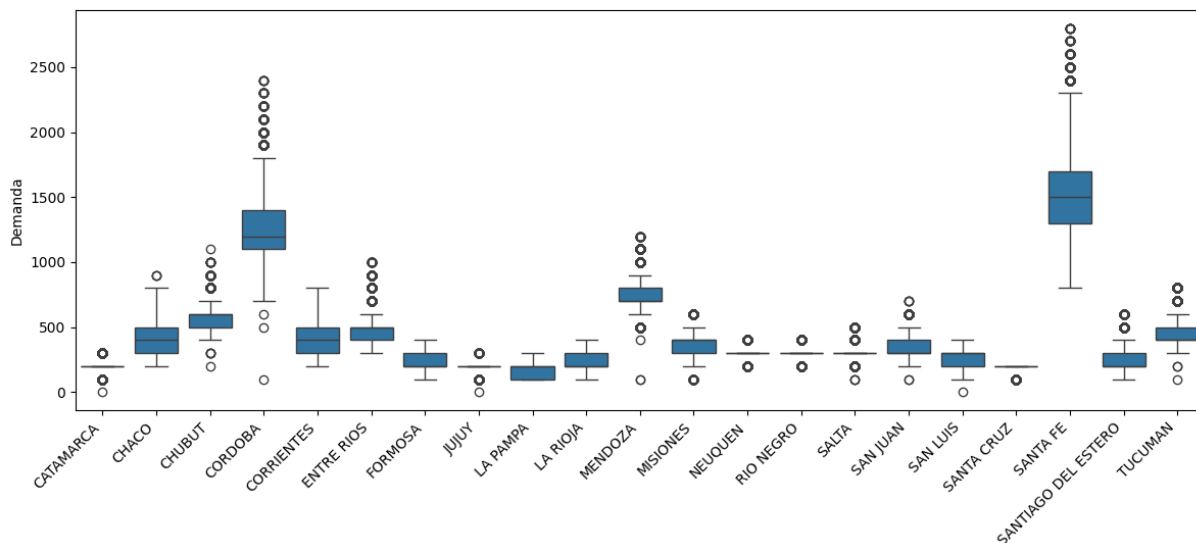


Gráfico 6 - Distribución de la demanda por provincia (excl. Bs. As.).

En el gráfico anterior se observa una marcada concentración de demanda en provincias como Santa Fe y Córdoba, mientras que otras jurisdicciones presentan menor rango intercuartílico y valores atípicos más acotados. La exclusión de Buenos Aires facilita resaltar estos contrastes.

A continuación, se muestran los boxplots generados agrupados por las variables *TipoDia* y *DiaSemana*, empleando la misma función `graficar_boxplot_demanda()`. Esto permite explorar diferencias en el patrón de consumo entre días hábiles, no hábiles y feriados, y entre días de semana y fines de semana.

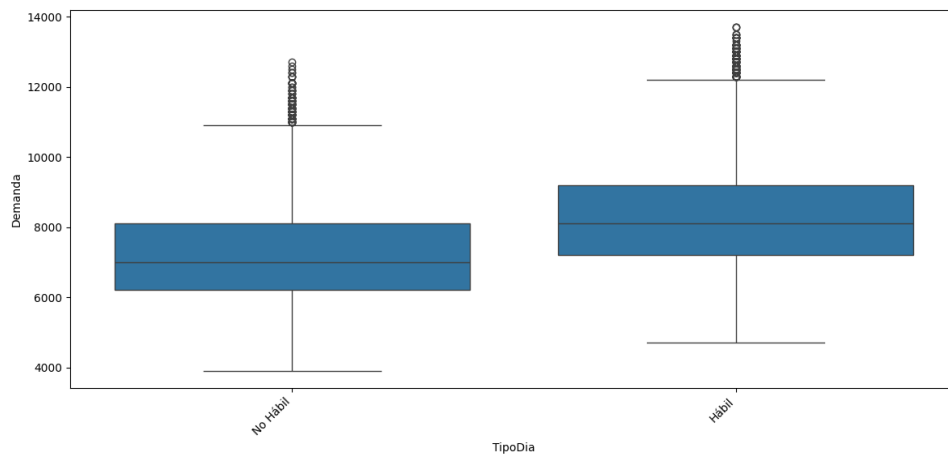


Gráfico 7 - Distribución de la demanda según el tipo de día.

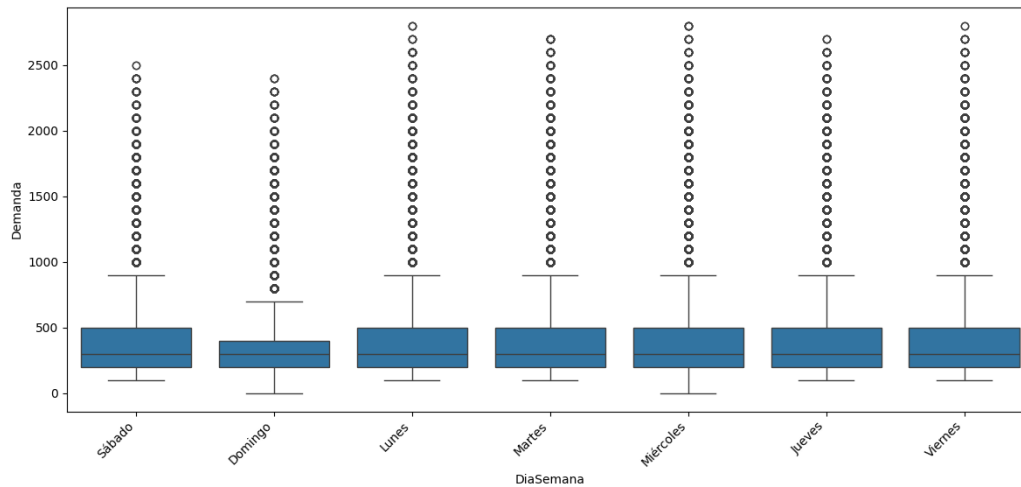


Gráfico 8 - Distribución de la demanda según el día de la semana.

El análisis horario de la demanda energética revela patrones de consumo claramente diferenciados según el tipo de día. Durante los días hábiles, se observa un perfil de demanda con

características marcadamente comerciales e industriales, iniciando con consumos relativamente bajos durante las primeras horas de la madrugada, seguido de un incremento sostenido que alcanza su máximo durante las horas de mayor actividad. En contraste, los días no hábiles presentan un patrón más uniforme y con niveles generalmente inferiores, manteniendo un consumo base más estable a lo largo del día, con picos menos pronunciados que reflejan principalmente el uso residencial. La diferencia más notable se presenta durante las horas centrales del día, donde la brecha entre días hábiles y no hábiles puede ser muy amplia, evidenciando el impacto significativo de la actividad comercial e industrial en el perfil de demanda.

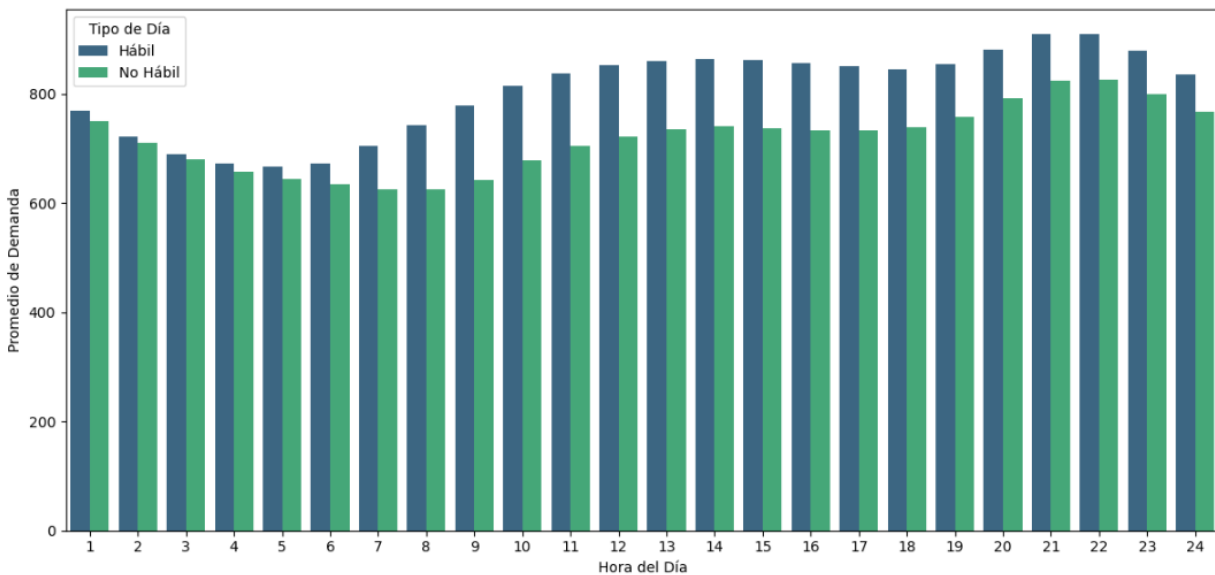


Gráfico 9 - Promedio de demanda por tipo de día y horario del día.

7.3.4 Conclusión del análisis exploratorio

Los gráficos generados muestran la existencia de estacionalidad, variabilidad horaria y diferencias por provincia y contexto calendárico, reforzando la hipótesis de que la demanda eléctrica está altamente condicionada por factores climáticos y horarios. Estos resultados preliminares justifican la selección de modelos de regresión robustos y redes neuronales profundas capaces de capturar relaciones no lineales y patrones complejos.

7.4 Preparación de features y preprocesamiento

El proceso de preparación de datos y preprocesamiento es fundamental para garantizar la calidad del entrenamiento de los modelos que serán entrenados y la reproducibilidad del Pipeline desarrollado .

7.4.1 Preparación de features

Con esta preparación se obtuvo un conjunto de entrenamiento con 462.932 registros con 10 variables independientes y un conjunto de prueba con 115.734 registros con las mismas variables. Esta separación balanceada garantiza que el modelo se entrene con una muestra robusta y se evalúe sobre un conjunto independiente.

7.4.2 Preprocesamiento: Codificación de variables categóricas para Pipeline

Se identifican las variables categóricas *Provincia*, *TipoDia*, *DiaSemana* y *EsFeriado*, que contienen categorías nominales que requieren ser transformadas a formato numérico mediante codificación one-hot. Se conservaron sin transformación las variables *Hora*, *TempC*, *Humedad*, *Precipitacion*, *Importacion* y *Exportacion*.

En la definición del preprocesador se definieron las siguientes configuraciones.

- *handle_unknown='ignore'* previene errores en predicción cuando aparecen categorías no vistas durante el entrenamiento.
- *remainder='passthrough'* asegura que las variables numéricas se mantengan sin modificar.

Con todo lo mencionado, los datos están listos para integrarse en un Pipeline de entrenamiento, asegurando que el preprocesamiento se aplique de manera consistente en todas las etapas de validación y despliegue.

7.5 Entrenamiento de modelos

Con los datos ya preprocesados y listos para el modelado, como se vio en las secciones anteriores, se procedió al entrenamiento de múltiples modelos de regresión con el objetivo de predecir la demanda eléctrica horaria en función de variables temporales, geográficas, meteorológicas y energéticas. Para lograr un enfoque comparativo robusto como se ha establecido, se entrenaron y compararon cuatro modelos distintos (Random Forest Regressor, Gradient Boosting Regressor, IRBDNN con bloques residuales y IRBDNN sin bloques residuales).

Todos los modelos fueron entrenados utilizando los conjuntos *X_train_processed* y *y_train*, y validados sobre *X_test_processed* y *y_test*.

7.5.1 Random Forest Regressor

Se implementó un modelo de tipo Random Forest (*RandomForestRegressor*), que construye múltiples árboles de decisión y promedia sus predicciones. Este enfoque es conocido por su capacidad

para manejar datos con alta dimensionalidad y por su bajo riesgo de sobreajuste cuando se configura correctamente. Se configuraron los siguientes parámetros.

- $n_estimators = 150$: cantidad de árboles.
- $max_depth = 12$: profundidad máxima de los árboles.
- $min_samples_split = 10$ y $min_samples_leaf = 5$: control del crecimiento de los árboles.
- $n_jobs = -1$: uso de todos los núcleos del procesador.

7.5.2 Gradient Boosting Regressor

Se entrenó un modelo GradientBoostingRegressor, un algoritmo de boosting basado en árboles que construye en cada pasada un nuevo árbol tratando de corregir los errores del paso anterior. Se configuraron los siguientes parámetros para mejorar el rendimiento y evitar el overfitting.

- $n_estimators = 300$, $learning_rate = 0.05$
- $max_depth = 6$, $subsample = 0.8$
- Regularización mediante $min_samples_split = 10$ y $min_samples_leaf = 5$

7.5.3 Red Neuronal IRBDNN con bloques residuales

Para capturar relaciones no lineales complejas entre las variables, se diseñó una red neuronal profunda inspirada en arquitecturas residuales. Este modelo, denominado IRBDNN Residual, incorpora conexiones tipo ResNet, lo cual facilita el flujo del gradiente y mejora la capacidad de generalización de redes profundas. En la aplicación de este modelo resultan particularmente relevantes las siguientes características.

Estructura

- Capa densa inicial con activación *ReLU*
- Tres bloques residuales (*res_block*) de la forma: $x \rightarrow Dense \rightarrow Dense \rightarrow Add(x) \rightarrow ReLU$
- Capa de salida lineal ($Dense(1)$)

Configuración del optimizador

- Adam con $learning_rate=0.0001$
- $loss='mse'$, métricas: MAE y MSE

Callbacks utilizados

- *EarlyStopping*: detiene el entrenamiento si val_loss no mejora en 10 épocas.

- *ReduceLROnPlateau*: reduce el learning rate si el modelo se estanca.

7.5.4 Red Neuronal IRBDNN Pura (sin bloques residuales)

Con fines comparativos, se entrenó una versión más simple del modelo anterior, sin bloques residuales utilizando los mismos hiperparametros, optimizador y callbacks. Esta red neuronal contiene:

- Capa de entrada + 3 capas ocultas densas con activación *ReLU*
- Capa de salida lineal

7.6 Evaluación de modelos

Finalizada la etapa de entrenamiento, se procedió a evaluar los modelos implementados mediante un procedimiento sistemático que permitió cuantificar el rendimiento predictivo y realizar una comparación objetiva entre ellos. Para dicha evaluación se desarrolló la función *evaluar_modelos()*; la cual cumple un doble propósito.

- Calcular métricas cuantitativas de desempeño sobre el conjunto de prueba (X_{test}, y_{test}).
- Generar visualizaciones del rendimiento del modelo en forma de gráficos Valor Predicho vs Valor Real para facilitar la inspección visual de los resultados.

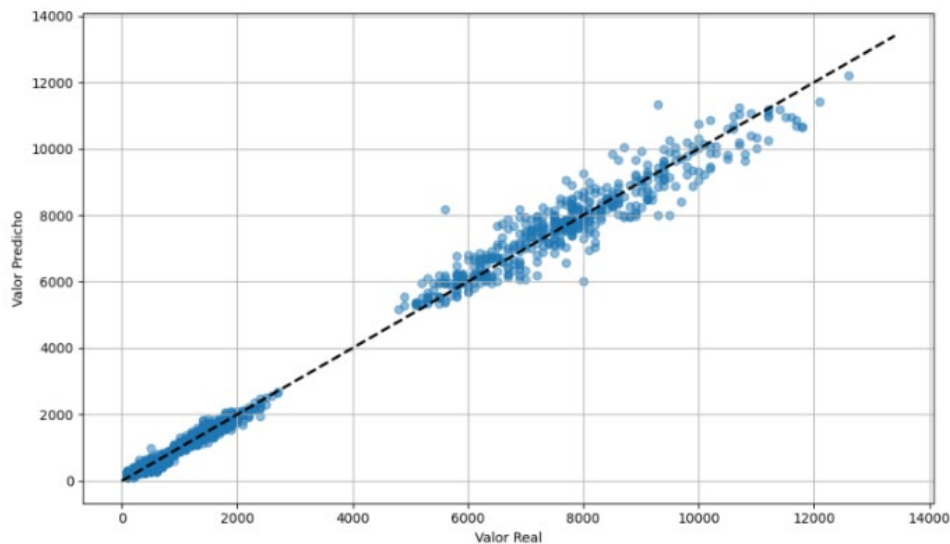


Gráfico 20 - Predicho vs Real - Modelo: RandomForest. Demuestra alta capacidad predictiva con excelente alineación a la línea de identidad, evidenciando la efectividad del promedio agregado entre estimadores.

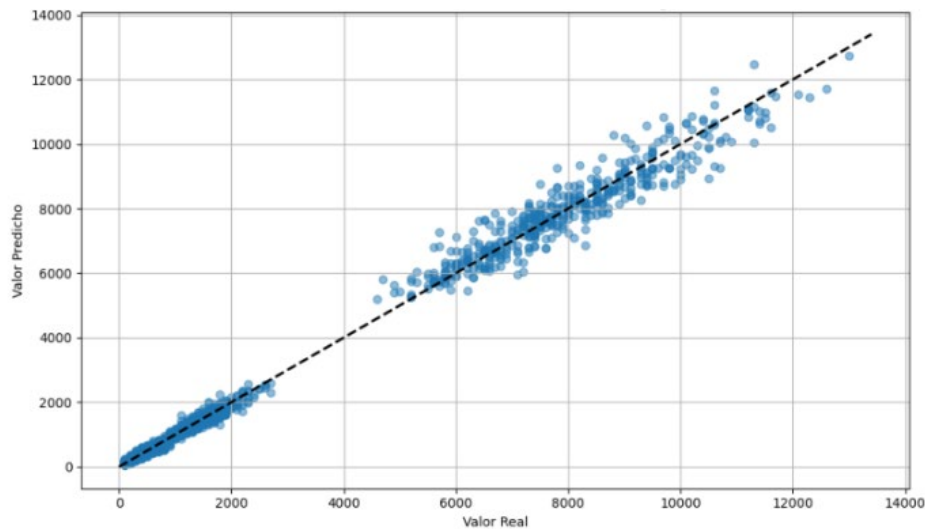


Gráfico 31 - Predicho vs Real - Modelo: Gradient Boosting. El algoritmo de boosting secuencial muestra excelente capacidad predictiva con puntos densamente agrupados alrededor de la línea de identidad, indicando predicciones consistentes y precisas.

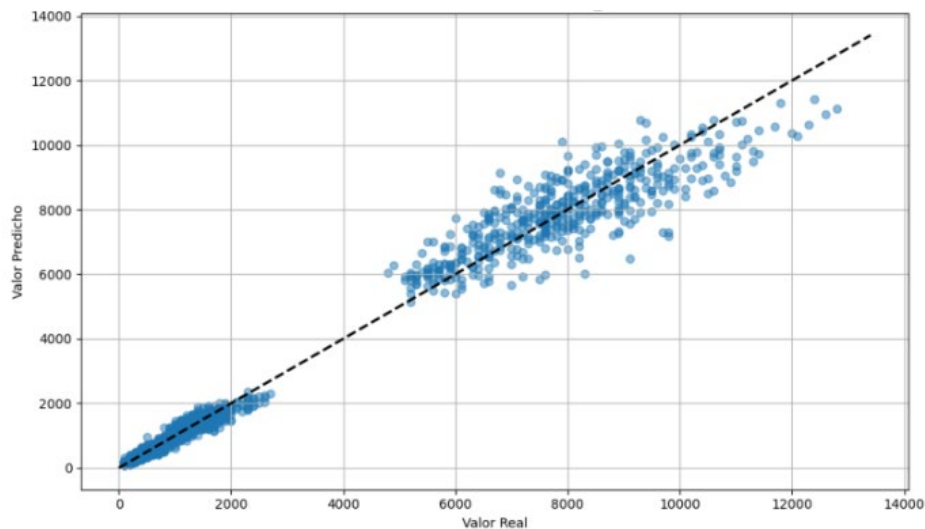


Gráfico 42 - Predicho vs Real - Modelo: IRBDNN_Residual. La arquitectura residual muestra un mejor flujo de gradientes, evidenciado por la alta correlación lineal entre valores observados y predichos, con mínima dispersión respecto a la línea de referencia en valores bajos de demanda.

Los resultados numéricos y los gráficos son guardados en el directorio especificado (*output_dir*). Las métricas utilizadas fueron las que se detallan a continuación, las mismas fueron calculadas con *mean_absolute_error*, *mean_squared_error* y *r2_score* de *sklearn.metrics*.

- MAE (Mean Absolute Error): error absoluto promedio
- MSE (Mean Squared Error): error cuadrático medio
- RMSE (Root Mean Squared Error): raíz cuadrada del MSE, más interpretable por estar en la misma escala que la variable objetivo

- R^2 (Coeficiente de determinación): medida de ajuste del modelo (1 = ajuste perfecto)

Los valores obtenidos para cada métrica fueron los que se detallan a continuación.

Modelo	MAE	MSE	RMSE	R^2
Random Forest	68.73	17,088.57	130.72	0.99
Gradient Boosting	53.11	14,609.97	120.87	0.99
IRBDNN Residual	67.81	24,728.87	157.25	0.99
IRBDNN Puro	80.20	44,091.72	209.98	0.98

Tabla 1 - Tabla comparativa de los modelos

Como se puede observar en la tabla anterior, el modelo Gradient Boosting presentó el mejor desempeño en términos de error (menor MAE y RMSE), con un ajuste excelente ($R^2 = 0.99$), posicionándose como la mejor opción general para este caso de estudio.

El modelo Random Forest también mostró un rendimiento sólido y estable, muy cercano al anterior; mientras que el modelo IRBDNN con bloques residuales logró un buen ajuste, aunque con mayor error absoluto y cuadrático.

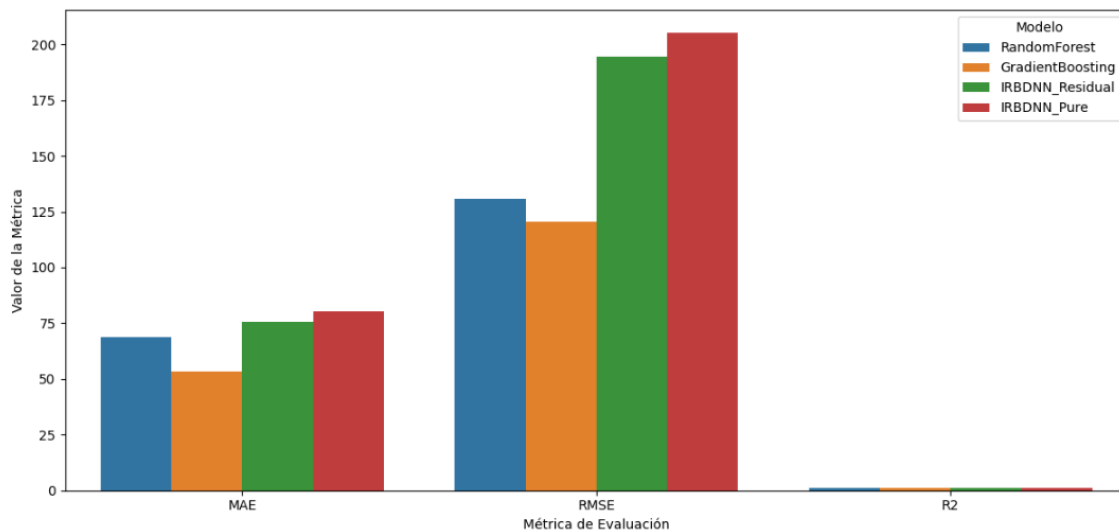


Gráfico 53 - Comparación de métricas de los Modelos. Análisis comparativo de performance mediante métricas MAE, RMSE y R^2 para los algoritmos implementados. El gráfico de barras permite evaluar simultáneamente las métricas, facilitando la selección del modelo óptimo basado en múltiples criterios de evaluación.

7.7 Selección del modelo final

A partir del análisis de métricas cuantitativas y la inspección visual de los resultados, el modelo Gradient Boosting Regressor se destaca como la mejor alternativa para la predicción de la demanda eléctrica en el contexto de este estudio.

7.7.1 Rendimiento cuantitativo

Como se observó en la tabla de métricas obtenidas en el apartado anterior, este modelo se destacó por obtener los siguientes valores

- El menor MAE, lo que implica que, en promedio, sus predicciones se alejaron menos de los valores reales.
- El menor RMSE, reflejando una baja dispersión de los errores incluso en valores extremos.
- Un coeficiente de determinación R^2 de 0.99, evidenciando un ajuste excelente a los datos observados.

7.7.2 Análisis visual

En el gráfico 11 se puede observar que la mayoría de los puntos se alinean fuertemente a lo largo de la diagonal que representa la predicción perfecta (valor real = valor predicho). Además de destacar que no se detectan sesgos sistemáticos ni zonas de sobreestimación o subestimación significativas.

La nube de puntos muestra un patrón estrecho y concentrado, incluso en valores altos de demanda, lo que indica consistencia y precisión en distintos rangos de consumo.

7.7.3 Robustez y generalización

El modelo Gradient Boosting mostró:

- Estabilidad durante el entrenamiento (sin overfitting¹³ significativo).
- Capacidad de generalización, manteniendo desempeño consistente frente a variaciones del conjunto de prueba.
- Habilidad para capturar relaciones no lineales y efectos de interacción entre variables, aspecto clave en un problema multifactorial como la demanda eléctrica, fuertemente influida por variables exógenas (clima, calendario y comportamiento del sistema eléctrico).

¹³ El overfitting, o sobreajuste, es un problema común en el aprendizaje automático donde un modelo se ajusta demasiado a los datos de entrenamiento, incluyendo ruido y patrones irrelevantes. Esto resulta en un alto rendimiento en el conjunto de entrenamiento, pero un rendimiento deficiente con datos nuevos o no vistos.

7.7.4 Comparación con otros modelos

Si bien todos los modelos implementados alcanzaron desempeños satisfactorios, se observaron diferencias relevantes:

- Random Forest Regressor obtuvo métricas competitivas y buena interpretación de variables, pero mostró mayor dispersión en predicciones de alta demanda, lo que puede limitar su uso en escenarios extremos.
- IRBDNN Residual logró un R^2 elevado y la ventaja de aprender patrones complejos, aunque su entrenamiento fue más costoso computacionalmente y la dispersión de errores (RMSE) resultó mayor respecto a Gradient Boosting.
- IRBDNN Puro presentó el menor desempeño relativo, con errores más altos y menor capacidad de ajuste, posiblemente debido a la ausencia de bloques residuales que mejoran la propagación de gradientes en redes profundas.

En este contexto, Gradient Boosting combina la interpretabilidad de los modelos de árbol con la capacidad de aprendizaje secuencial y corrección iterativa, lo que le permite alcanzar un balance óptimo entre precisión, robustez y eficiencia computacional.

7.7.5 Justificación de la elección

Teniendo en cuenta lo explicado con anterioridad

- Su alto poder explicativo y estabilidad.
- Su bajo nivel de error absoluto y relativo.
- Su desempeño consistente en el análisis visual y cuantitativo.

Se concluye que el modelo Gradient Boosting Regressor es el más adecuado para ser considerado como modelo final en esta etapa del trabajo, con el Dataset utilizado para el análisis.

Este modelo será utilizado en etapas posteriores para:

- Realizar predicciones sobre nuevos datos.
- Evaluar el impacto de variables críticas como el horario, tipo de día, feriados y condiciones climáticas.
- Proyectar escenarios futuros de demanda eléctrica en Argentina, contribuyendo a la planificación energética con herramientas de Ciencia de Datos reproducibles y escalables.

7.8 Interpretación del Modelo: Importancia de las Variables

Se realizó una evaluación de la importancia relativa de cada variable predictora.

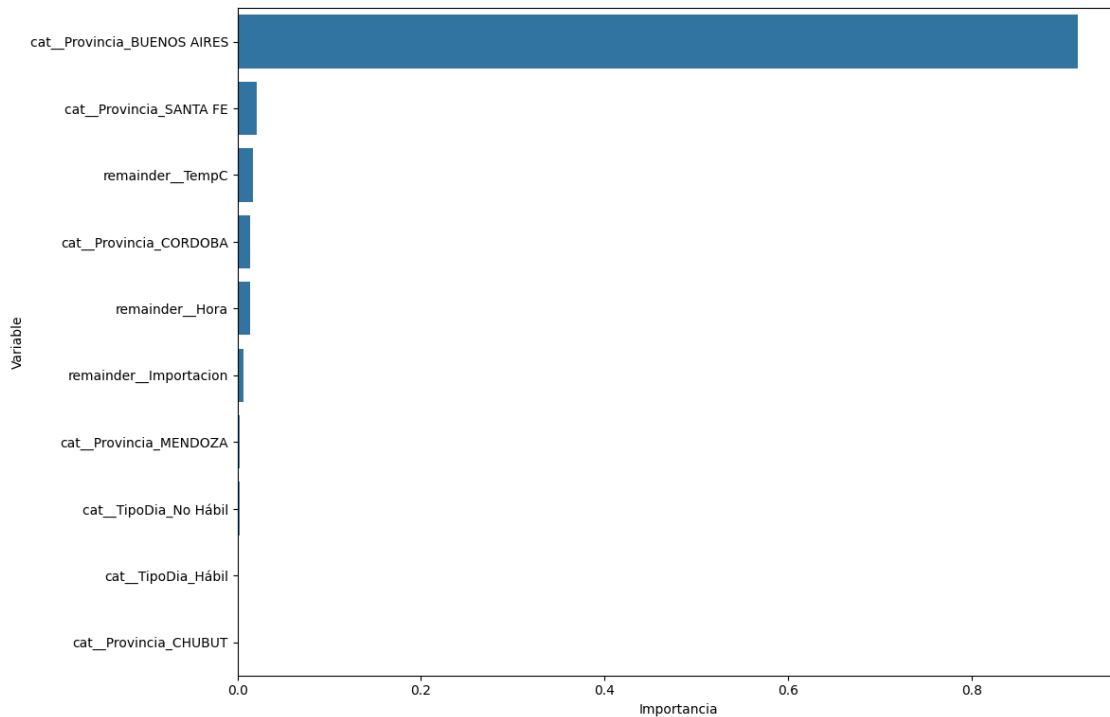


Gráfico 64 - Top 10 de variables predictoras.

Otras provincias como Santa Fe, Córdoba, Mendoza y Chubut aparecen con pesos significativamente menores, pero relevantes dentro del conjunto de datos. Entre las variables exógenas al modelo se priorizan las siguientes.

- *TempC*: influye con la estacionalidad del consumo eléctrico por calefacción y refrigeración.
- *Hora*: refleja los patrones de comportamientos típicos de consumo durante el día.
- *Importación*: aporta contexto sobre el estado del sistema eléctrico.

Las variables relacionadas al tipo de día (*TipoDia*) también tienen impacto, aunque en menor proporción, reafirmando los patrones detectados en los análisis exploratorios previos que mostraban mayores consumos en días laborables.

Capítulo 8. Futuros estudios

La fuerte concentración en la variable categórica de Provincia de Buenos Aires podría sugerir que el modelo puede estar sesgado hacia esta provincia. Podría ser de gran aporte futuros análisis e

implementación de modelos diferenciados por región o aplicar técnicas de balance. Los resultados alcanzados en este trabajo ponen de manifiesto la necesidad de continuar profundizando en el análisis de la demanda eléctrica en Argentina.

Otra línea de trabajo futura es la incorporación de escenarios prospectivos que contemplen factores emergentes como la penetración de energías renovables, la movilidad eléctrica o las olas de calor intensificadas por el cambio climático. Estos escenarios permitirían brindar insumos estratégicos para la planificación energética de mediano y largo plazo.

De esta forma, el presente estudio sienta las bases de un pipeline metodológico sólido y escalable, al tiempo que abre nuevas oportunidades de investigación aplicada para enriquecer la toma de decisiones en el sector energético argentino.

Capítulo 9. Conclusiones y recomendaciones

9.1 Conclusión

El presente trabajo permitió implementar y comparar distintos enfoques de modelado predictivo sobre la demanda eléctrica horaria en Argentina, utilizando datos reales del período 2022–2024 e incorporando variables exógenas como clima, tipo de día e intercambios de energía. A través del desarrollo de distintos modelos. Se logró alcanzar un alto nivel de precisión, destacándose particularmente el rendimiento del modelo Gradient Boosting.

El análisis de la importancia de las variables y la segmentación geográfica revelaron patrones clave del comportamiento energético, como el peso significativo de la provincia de Buenos Aires, la influencia del horario y la temperatura, y las diferencias entre días hábiles y no hábiles. Estos hallazgos refuerzan la utilidad de la ciencia de datos para mejorar la planificación energética nacional y regional.

9.2 Recomendaciones prácticas

A partir de los resultados obtenidos en este trabajo, se proponen las siguientes recomendaciones prácticas:

Implementar herramientas de predicción de demanda en tiempo real a nivel nacional y regional, utilizando el modelo de Gradient Boosting, que demostraron alto rendimiento en relación con los otros modelos analizados; lo que permitirá poder anticipar picos de demanda y gestionar eficientemente los recursos.

Incorporar variables climáticas en la planificación energética futura, ya que se evidenció su impacto significativo sobre el consumo eléctrico. Desarrollar Dashboards interactivos para monitorear la demanda horaria, clasificada por provincia, tipo de día y eventos especiales como feriados, permitiendo de esta forma un monitoreo más accesible para operadores y analistas.

Revisar los mecanismos de importación/exportación energética, ajustándolos según predicciones de consumo, evitando desbalances que generen fallas en el SADI.

Aplicar modelos específicos por provincia o clusterizar, dada la fuerte heterogeneidad del comportamiento energético según región. Extender el uso del modelo a predicciones futuras con distintos escenarios climáticos y de crecimiento poblacional.

Capítulo 10. Apéndice técnico

Repositorio de lo desarrollado en este apartado

https://drive.google.com/drive/folders/1TEzK5KdI2YzA7W5TdFK8Cnft-krBipXi?usp=drive_link

10.1 Código fuente

La solución fue desarrollada en Python, implementada y validada en el entorno Google Colab, y organizada en bloques temáticos para facilitar su mantenimiento y reutilización. Utilizando las siguientes librerías pandas, numpy, scikit-learn, xgboost, openpyxl, requests, seaborn y matplotlib.

10.1.1 Preparación de los Dataset y ejecución

- `Genera_demanda_desnormalizada.ipynb`: transforma el archivo Excel de demanda por provincia de formato ancho a formato largo, filtra solo los años 2022, 2023 y 2024 redondeando los valores de la demanda hacia arriba al múltiplo de 100 más cercano.
- `Genera_importacion_exportacion_2022_2024.ipynb`: transforma el archivo Excel de importación/exportación horaria a CSV filtrando datos hasta el 31/12/2024.
- `Genera_feridos_argentina_2022_2024.ipynb`: consulta la API Nager.Date para obtener feriados por año.
- `Genera_Clima_Argentina_2022_2024.ipynb`: consulta la API Open-Meteo para cada provincia y hora, generando un Dataset final con las variables temperatura, humedad relativa y precipitación. Preprocesamiento, entrenamiento, evaluación y análisis
- `Integrador_Prediccion_Demanda.ipynb`: este archivo se compone de los siguientes módulos.

- a. Librerías necesarias.
- b. Carga y unificación de datos.
- c. Análisis exploratorio.
- d. Preparación de features
- e. Preprocesamiento: Codificación de variables categóricas para Pipeline
- f. Entrenamiento de modelos
- g. Evaluación de modelos
- h. Visualización de resultados

10.2 Fuentes de Datos y Datasets Utilizados

Todos los datos fueron almacenados en formato CSV estandarizado y utilizados sin modificaciones externas, salvo por procesos de limpieza, transformación de fechas, normalización horaria y codificación para modelado.

- demanda_por_provincia.csv
- demanda_por_provincia_2022_2024.csv
- Import-Export Horaria.xlsx
- importacion_exportacion_2022_2024.csv
- feriados_argentina_2022_2024.csv
- clima_argentina_2022_2024.csv
- Combinado_2022_2024.csv

Dataset	Fuente	Enlace
Demanda horaria por provincia (2022–2024)	CAMMESA	https://www.cammesa.com/
Importación / exportación horaria	CAMMESA	https://www.cammesa.com/
Datos climáticos por hora y provincia	Open-Meteo API	https://open-meteo.com/
Feridos nacionales	Nager.Date API	https://date.nager.at/Api

Tabla 2 - Datasets principales utilizados

Referencias bibliográficas

- Abadi, M. A. (s.f.). *TensorFlow: Large-scale machine learning on heterogeneous systems*. (Version 2.x) [Software]. Obtenido de <https://www.tensorflow.org/>
- Chollet, F. &. (2015). *Keras: The Python Deep Learning API*. (Version 2.x) [Software]. Obtenido de <https://keras.io/>
- Developers, N. (2023). NumPy. (Version 1.x) [Software]. Obtenido de <https://numpy.org/>
- Hafeez, G. A. (2020). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Applied Energy*, 269, 114915. doi:10.1016/j.apenergy.2020.114915
- Hammad, M. A. (2020). Methods and models for electric load forecasting: A comprehensive review. *Logistics & Sustainable Transport*, 11(1), 51–76. doi:10.2478/jlst-2020-0004
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. Obtenido de <https://matplotlib.org/>
- McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445, 51–56. Obtenido de <https://pandas.pydata.org/>
- Nti, I. K.-B. (2020). Electricity load forecasting: a systematic review. *Journal of Electrical Systems and Information Technology*, 7(1), 13. doi:10.1186/s43067-020-00021-8
- Pedregosa, F. V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. Obtenido de <https://scikit-learn.org/>
- Rostum, M. A. (2020). Electrical Load Forecasting: A methodological overview. *Ministry of Electricity & Energy, Egypt*. (Z. University, Ed.)
- Solyali, D. (2020). A Comparative Analysis of Machine Learning Approaches for Short-/Long-Term Electricity Load Forecasting in Cyprus. *Sustainability*, 12(9), 12(9), 3612. doi:10.3390/su12093612
- Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. doi:10.21105/joss.03021