Power demand forecasting of university buildings with Random Forest and Stacking algorithms

Maza Díaz, Yenia[1], Funes, Marcos[1], Donato, Patricio[1], and Orallo, Carlos[1]

Instituto de Investigaciones Científicas y Tecnológicas en Electrónica (ICYTE), UNMDP-CONICET, Mar del Plata, Argentina yenialuna@gmail.com

Abstract Electric load forecasting in institutional environments presents specific challenges due to the irregular consumption patterns, particularly during academic recess periods or after structural changes in electrical infrastructure. This work evaluates the forecasting performance of Random Forest and Stacking models using load data from two buildings at the National University of Mar del Plata. A data augmentation strategy based on controlled-amplitude Gaussian noise is assessed to improve model generalization. The results show that incorporating Gaussian noise reduces the maximum error excursion of the Stacking model (by at least 50% in the extreme values for FCEyS and below a 15% maximum excursion for FI), at the cost of a slight increase in the mean error. The Random Forest algorithm does not exhibit the same favorable response to noise, suggesting that the effectiveness of augmentation depends on the type of model and the underlying demand profile.

Keywords: Smart grids , Power demand , Forecasting , Random Forest , Stacking.

Pronóstico de demanda eléctrica de edificios universitarios mediante algoritmos Random Forest y Stacking

Resumen El pronóstico de la demanda eléctrica en entornos institucionales presenta desafíos particulares derivados de la irregularidad de los patrones de consumo, especialmente durante períodos de receso académico o ante cambios estructurales en la infraestructura eléctrica. En este trabajo se evalúa la capacidad de pronóstico de los algoritmos Random Forest y Stacking, usando datos de demanda de la Universidad Nacional de Mar del Plata. Adicionalmente, con el objetivo de mejorar la capacidad de generalización de los modelos de aprendizaje automático, se evalúa la estrategia de aumento de datos basada en la adición de ruido gaussiano de amplitud controlada al conjunto de entrenamiento. Los resultados obtenidos muestran que se reduce la excursión máxima del error del modelo Stacking (al menos un 50 % en los valores extremos para FCEyS y por debajo de un 15 % de excursión máxima para FI), a costa de un leve incremento del error promedio. El algoritmo Random Forest no exhibe la misma sensibilidad positiva al ruido, lo que sugiere que la efectividad de la estrategia de aumento de datos depende del tipo de modelo y del perfil de demanda.

Keywords: Redes eléctricas inteligentes , Demanda eléctrica , Pronósticos , Random Forest , Stacking.

1. INTRODUCCIÓN

El crecimiento sostenido de la demanda eléctrica global, impulsado por la expansión urbana y la proliferación de nuevas cargas tecnológicas, ha posicionado al pronóstico de corto plazo como una herramienta estratégica en la gestión de las Redes Eléctricas Inteligentes (REI) (International Energy Agency [IEA], 2025; Powell et al., 2024). En este contexto, las técnicas de aprendizaje automático han ganado protagonismo por su capacidad para modelar relaciones no lineales y patrones temporales complejos en series de tiempo eléctricas (De Caro et al., 2020; Mahmood et al., 2024).

En un trabajo previo (Maza Diaz et al., 2024), se evaluó el desempeño de cinco algoritmos de aprendizaje automático para el pronóstico de demanda eléctrica en un edificio de la Universidad Nacional de Mar del Plata (UNMDP). El pronóstico de demanda eléctrica en edificios universitarios constituye una herramienta clave para la gestión energética institucional ya que permite optimizar la operación de las instalaciones, anticipar picos de consumo, reducir costos energéticos y favorece la integración de fuentes renovables. La presencia de generación fotovoltaica en los edificios analizados convierte al pronóstico de demanda en una herramienta fundamental para la gestión del almacenamiento energético, el autoconsumo y la interacción con la red eléctrica. En trabajos preliminares se concluyó que Random Forest y Stacking exhiben un mayor equilibrio entre precisión y generalización respecto de alternativas como K-Nearest Neighbors (KNN), Ridge, Lasso o Gradient Boosting Trees. Posteriormente se extendió dicho análisis aplicando el algoritmo RF a cuatro edificios con perfiles heterogéneos: la Facultad de Ciencias Exactas (FCE), la Facultad de Ingeniería (FI), la Facultad de Ciencias Económicas y Sociales (FCEyS) y el Instituto de Investigaciones en Ciencia y Tecnología de Materiales (INTEMA) (Maza Diaz et al., 2025). Los resultados revelaron que el pronóstico se degrada cuando la demanda cae fuera del rango histórico de entrenamiento, en particular durante los períodos de receso académico y frente a cambios estructurales en la infraestructura eléctrica. En particular, se observó que, por diferentes motivos los pronósticos más afectados fueron los de FI y FCEyS.

Este trabajo tiene como objetivo extender estudios previos aplicando el algoritmo Stacking (ST) a las series de datos de demanda de los dos edificios con los perfiles de demanda más desafiantes, FI y FCEyS; comparar su desempeño con el de Random Forest (RF) bajo condiciones equivalentes; y evaluar el efecto de una estrategia de aumento de datos (del inglés *data augmentation*) basada en ruido gaussiano sobre la capacidad de generalización de ambos modelos. El principal aporte no radica en la técnica de aumento en sí, ampliamente documentada, sino en la caracterización empírica de su efectividad en función del modelo y del perfil de demanda.

2. Modelos de Pronóstico

Los modelos de aprendizaje automático supervisado han demostrado una capacidad notable para capturar patrones no lineales en series temporales de

demanda eléctrica (Deb et al., 2017). En este enfoque, la serie se representa mediante vectores construidos a partir de rezagos temporales, estadísticas de ventana y variables calendario, permitiendo modelar patrones complejos sin asumir una forma funcional paramétrica para el proceso generador de datos. Esto los hace especialmente adecuados para sistemas con dinámicas complejas, como los edificios universitarios, donde la demanda responde simultáneamente a horarios académicos, ciclos estacionales, ocupación edilicia y eventos extraordinarios (Deb et al., 2017). El proceso de entrenamiento supervisado consiste en ajustar una función que minimice el riesgo empírico sobre la serie a predecir. La generalización del modelo depende de que la distribución de los datos de entrenamiento sea representativa de la distribución del conjunto de test. Cuando este supuesto falla, como ocurre durante los recesos académicos o ante modificaciones en la infraestructura, la capacidad de generalización puede deteriorarse significativamente. La elección de RF y ST está basada en la evidencia empírica previa (Maza Diaz et al., 2024) y en su adecuación al tipo de datos tabulares con ingeniería de atributos como los aquí empleados. El método ST basado en árboles de decisión iguala y supera con frecuencia a las arquitecturas de aprendizaje profundo, con menor costo computacional y mayor interpretabilidad.

Random Forest (RF): Este algoritmo construye un conjunto de árboles de decisión entrenados de manera independiente a partir de distintas submuestras aleatorias de los datos originales, seleccionando además en cada nodo un subconjunto aleatorio de variables candidatas para la partición (Cutler et al., 2012). Esta doble aleatoriedad, tanto en los datos como en las variables, reduce la correlación entre los árboles y mejora la capacidad de generalización del modelo. Para realizar el pronóstico, cada árbol produce su propio resultado y el modelo combina dichas salidas. Si el problema es de clasificación, se le asigna la clase más votada, mientras que en problemas de regresión se calcula el promedio de las predicciones. Es ampliamente utilizado debido a su capacidad para manejar conjuntos de datos de alta dimensionalidad, su robustez al ruido y su interpretabilidad relativamente alta en comparación con otros métodos de aprendizaje automático. Sin embargo, para alcanzar un rendimiento óptimo es fundamental ajustar adecuadamente sus hiperparámetros, idealmente mediante técnicas como la validación cruzada.

Stacking (ST): Este algoritmo combina múltiples modelos individuales para mejorar el rendimiento predictivo general (Breiman, 1996). Se basa en la premisa de que diferentes modelos pueden identificar patrones complementarios en los datos y que, al combinarlos de manera adecuada, se pueden obtener predicciones más precisas que las que se obtienen con cualquiera de los modelos por separado. Para realizar un pronóstico, los modelos base se aplican primero a los nuevos datos y generan sus respectivas salidas. Estas predicciones se utilizan luego como entrada para un modelo de nivel superior, encargado de combinarlas y producir el pronóstico final. Stacking puede ser especialmente efectivo cuando los modelos base capturan diferentes aspectos del problema y sus errores no están correlacionados. Al combinar sus predicciones, el modelo puede aprovechar las fortalezas individuales de cada modelo base y compensar sus debilidades. Sin embargo,

Stacking también tiene algunas desventajas, como un mayor costo computacional debido a la necesidad de entrenar múltiples modelos, y la posibilidad de sobreajuste si no se implementa correctamente.

2.1. Aumento de datos en series temporales

El aumento de datos consiste en generar nuevas muestras de entrenamiento mediante transformaciones aplicadas a los datos originales, para incrementar la diversidad del conjunto de entrenamiento y mejorar la capacidad de generalización del modelo (Shorten y Khoshgoftaar, 2019). Esta estrategia permite reducir el sobreajuste al exponer al modelo a variaciones plausibles de los datos observados durante el proceso de aprendizaje. Si bien el aumento de datos ha sido ampliamente utilizado en otras áreas, su aplicación en el análisis de series temporales ha cobrado un interés creciente en los últimos años. Esto se debe a la limitada disponibilidad de datos etiquetados y a la necesidad de mejorar la robustez de los modelos frente a perturbaciones, ruido y variabilidad inherente a los sistemas dinámicos (Wen et al., 2021). Las técnicas de aumento incluyen la adición de ruido controlado, deformaciones temporales (time warping), recorte y deformación de ventanas (window slicing/warping), transformaciones en el dominio de frecuencia, Mixup y generación sintética mediante modelos generativos y GANs (Wen et al., 2021; Zhang et al., 2018). Entre estas, se adoptó la adición de ruido gaussiano por ser la de interpretación más directa y menor costo computacional, y por preservar las propiedades estadísticas de la señal mientras incrementa la variabilidad del conjunto de entrenamiento (Salamon y Bello, 2017). La estrategia consiste en añadir perturbaciones gaussianas a cada muestra de la serie de entrenamiento, manteniendo inalterado el conjunto de test. Para ello, se define un parámetro de escala (α) que controla la amplitud del ruido en relación a la varianza de la serie. Esta parametrización asegura que la perturbación sea proporcional a la amplitud de la señal, evitando introducir valores físicamente inconsistentes en la serie temporal (Salamon y Bello, 2017). Se adoptó $\alpha = 0{,}10$, mediante búsqueda en grilla sobre $\alpha \in \{0{,}05, 0{,}10, 0{,}20, 0{,}30, 0{,}50\}$ evaluando el MAE/RMSE en los datos de validación.

3. Descripción general de las bases de datos

Las bases de datos empleadas en este estudio corresponden a las series temporales de potencia activa de FI y FCEyS desde el 01/2022 al 10/2025 con una frecuencia de muestreo de 15 minutos, lo que resulta en 134,305 muestras por edificio. Las principales estadísticas descriptivas se presentan en la Tabla 1 y la Fig. 1 muestra la evolución temporal de estas variables. En la figura se ilustran en color naranja los segmentos utilizados para entrenamiento, y en color negro los segmentos utilizados para evaluación.

El preprocesamiento de las series temporales sigue el siguiente procedimiento. A partir de la serie de potencia activa con resolución de 15 minutos, se construyen vectores de características que incluyen: (a) rezagos temporales de la propia serie

Tabla 1: Magnitudes estadísticas de las bases de datos evaluadas.

Demanda	P_{max} (kW)	P_{med} (kW)	σ (kW)
FI	104,9	23,5	14,8
FCEyS	180,8	50,7	21,1

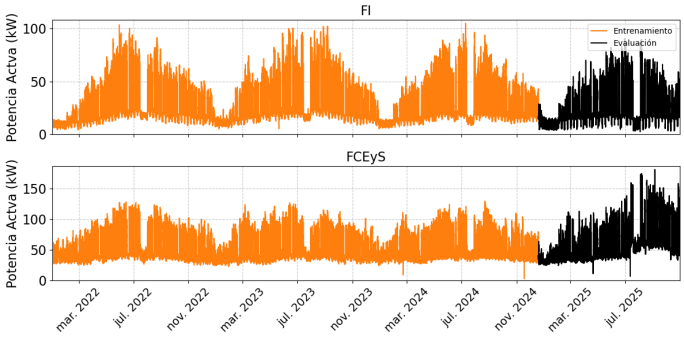

Figura 1: Potencia activa por base de datos.

(lag features) en instantes t-1, t-4, t-96 y t-672, correspondientes a 15 minutos, 1 hora, 24 horas y 7 días previos respectivamente; (b) estadísticos de ventana deslizante (media y desviación estándar en ventanas de 4, 96 y 672 muestras); y (c) variables calendario codificadas (hora del día, día de la semana, mes del año, indicador de día hábil). Los valores faltantes producto de interrupciones del servicio eléctrico fueron detectados e imputados mediante interpolación lineal antes del proceso de entrenamiento. Cabe destacar que durante el período analizado se produjeron dos eventos que alteraron las condiciones operativas de los edificios. A fines de 2024 se amplió la planta de generación fotovoltaica de FI, reduciendo la demanda neta en períodos de baja actividad, y durante la primera mitad de 2025 se modificó la topología eléctrica de FCEyS, derivando carga adicional desde FCE. Estos eventos, no representados en los datos de entrenamiento, constituyen un desafío para los modelos de pronóstico utilizados.

4. Resultados de pronóstico

En esta sección se analizan los resultados de las predicciones realizadas con los modelos en el periodo de evaluación comprendido entre el 01/11 y el 31/10/2025. Los modelos se implementaron en Python utilizando la biblioteca Scikit-Learn (Pedregosa et al., 2011). El modelo ST incluye tres estimadores de nivel base: un regresor XGBoost, un regresor LightGBM y un regresor Random Forest, cuyas predicciones son combinadas por una regresión lineal de nivel superior. Los hiperparámetros de cada estimador fueron fijados en los valores óptimos determinados en trabajos previos. Para el entrenamiento del meta-modelo, las predicciones de los estimadores base se generaron mediante validación cruzada

temporal con $k = 5$ particiones, obteniendo predicciones out-of-fold y evitando la validación cruzada aleatoria, inapropiada para series temporales. La partición global entrenamiento/prueba es temporal y el conjunto de prueba no interviene en ningún ajuste de hiperparámetros ni en la generación de meta-características.

Los errores se procesaron en dos segmentos temporales, S1 abarca los periodos de actividad institucional (01/11 al 31/12/2024, 03/02 al 18/07/2025, 04/08 al 31/10/2025) y S2 contiene los periodos de receso (01/01 al 02/02/2025, 19/07 al 03/08/2025). Se evaluaron por separado cada uno, ya que la demanda se reduce notoriamente en todas las instalaciones debido a la ausencia de personal en los periodos de receso (verano, invierno). Los errores se analizarán estadísticamente mediante diagramas de caja y se compararán los resultados.

4.1. Base FI

En la Fig. 2 se muestran los resultados de las predicciones realizadas sobre la base FI.

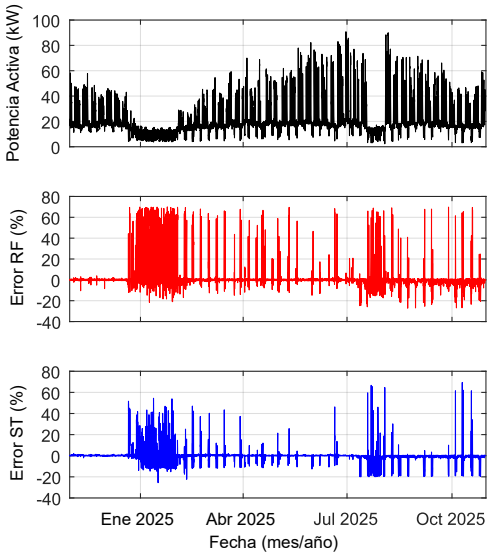

Figura 2: Base FI y errores relativos con RF y ST.

Se puede observar que tanto con RF como con ST se cometen errores superiores a un 40 % en los periodos de receso como en el resto de la demanda de 2025. En particular, con RF los errores son mayores que con ST. Durante los periodos de receso se observa que el error alcanza el 60 % de manera sistemática. Los resultados también revelan que, a lo largo del período de actividad de 2025, la magnitud del error se incrementa durante los fines de semana y días feriados, periodos caracterizados por una demanda mínima. Con ST, el error tiene una magnitud menor, pero con un patrón similar.

La Fig. 3 muestra el diagrama de caja de cada segmento de error. En esta figura se puede apreciar que el valor medio del error es esencialmente nulo para el

segmento S1 ($\overline{S1}_{RF} = -0,3446\,\%$, $\overline{S1}_{ST} = -0,0120\,\%$), y el 50 % de los valores es cercano a cero. Sin embargo se puede ver que con RF hay errores máximos en exceso de hasta aproximadamente un 70 % y en defecto de 25 %. Con ST, la excursión del error llega a valores correspondientes a los fines de semana y días feriados. Como se mencionó anteriormente, en periodos de receso, segmento S2, el comportamiento es similar respecto del segmento S1 con ambos algoritmos.

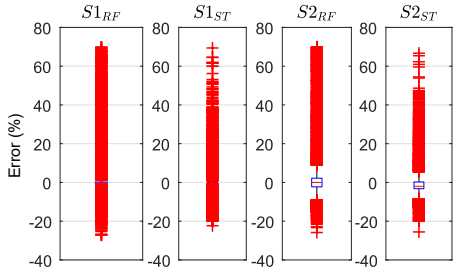

Figura 3: Base FI. Diagrama de caja del error cada segmento y algoritmo.

Respecto de los errores de pronóstico, las mayores desviaciones se concentran durante fines de semana y períodos de baja actividad académica. En estas condiciones, la demanda eléctrica disminuye significativamente y la generación fotovoltaica inyecta energía en la red interna, reduciendo aún más la demanda neta del edificio. En consecuencia, durante los días en los que no hay actividad, la demanda cae por debajo de aquellos valores de años anteriores. Al igual que en el caso de la base FCEyS, el algoritmo es susceptible de generar predicciones erróneas cuando la demanda eléctrica cae por debajo de los mínimos históricos con los que fue entrenado. La Fig. 4 muestra un ejemplo durante semana de vacaciones de invierno en 2025. La demanda de potencia cae hasta valores tan bajos como $4kW$ debido a que en las horas de sol la inyección de potencia de los paneles solares alcanza para cubrir casi toda la demanda del edificio. Ese fenómeno no se observó en los años previos a raíz de que la planta fotovoltaica era de menor potencia e impactaba menos en la demanda base de la instalación.

4.2. Base FCEyS

En la Fig. 5 se muestran los resultados de las predicciones realizadas sobre la base FCEyS. Se puede observar que tanto con RF como con ST desde finales de 2024 hasta cerca de julio de 2025 el desempeño de los algoritmos es muy bueno. El error es muy bajo, con algún incremento durante una parte del receso de verano. De hecho, se observa que con esta base no se producen errores significativos durante el receso de 2025. Se evidencia no obstante que desde julio en adelante se producen errores en defecto durante el periodo de actividad. A partir de la Fig. 6 se puede apreciar que el valor medio del error es esencialmente nulo para el segmento S1 ($\overline{S1}_{RF} = -0,3522\,\%$, $\overline{S1}_{ST} = -0,4930\,\%$), y el 50 % de los valores es cercano a cero. Sin embargo se puede ver que con RF hay errores máximos en

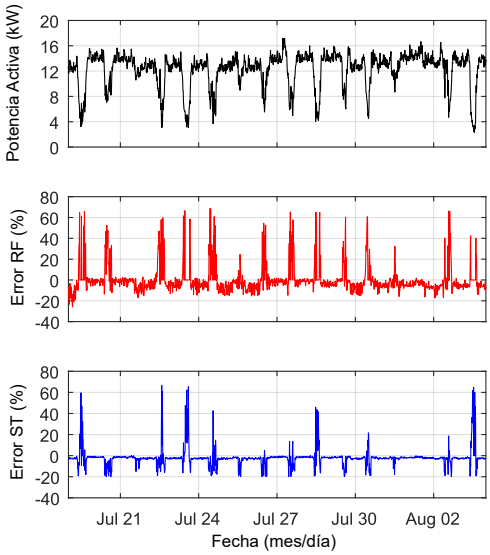

Figura 4: Base FI, receso de julio de 2025.

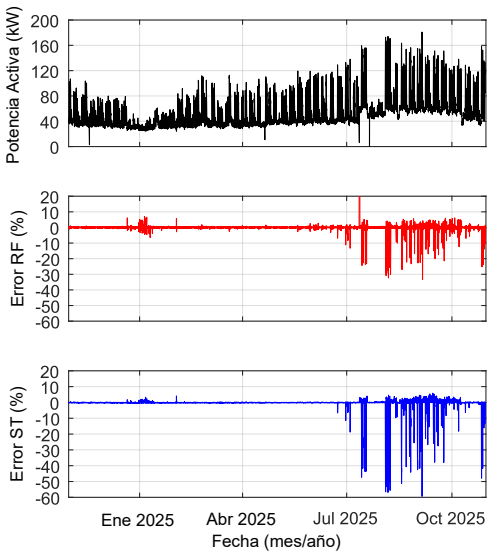

Figura 5: Base FCEyS (negro) y errores relativos con RF y ST.

defecto de hasta aproximadamente un 30 % y con ST de hasta un 60 %. Luego, durante S2, los errores son muy acotados con una mayor dispersión con RF.

La Fig. 7 presenta un extracto de la demanda de potencia y los errores de pronóstico correspondientes a agosto de 2025, evidenciando que ambos métodos subestiman de manera significativa la demanda máxima.

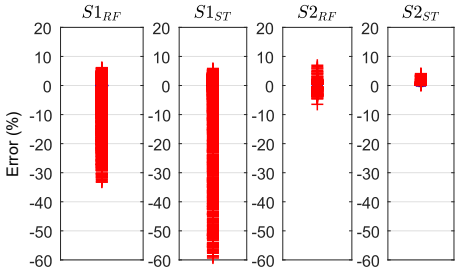

Figura 6: Base FCEyS. Diagrama de caja del error cada segmento y algoritmo.

Analizando este caso con respecto a los datos de entrenamiento, se observa que la demanda en el periodo 2022-2024 nunca alcanzó esos valores máximos y en consecuencia, los algoritmos de pronóstico no pueden predecir estos valores. Este incremento por encima de los valores históricos se explica por un cambio en la topología de la instalación eléctrica del edificio FCEyS. Durante la primera mitad de 2025, la demanda del edificio FCE alcanzó niveles muy elevados, que afectaban al transformador de distribución local. Por tal motivo, se modificaron los tableros de distribución, derivando parte de la carga al tablero de FCEyS. Este incremento en consecuencia no estaba incluido en los datos de entrenamiento.

Los resultados evidencian que el desempeño de los modelos de pronóstico analizados depende fuertemente de las características particulares de cada base de datos, especialmente en lo relativo a la estabilidad del comportamiento his-

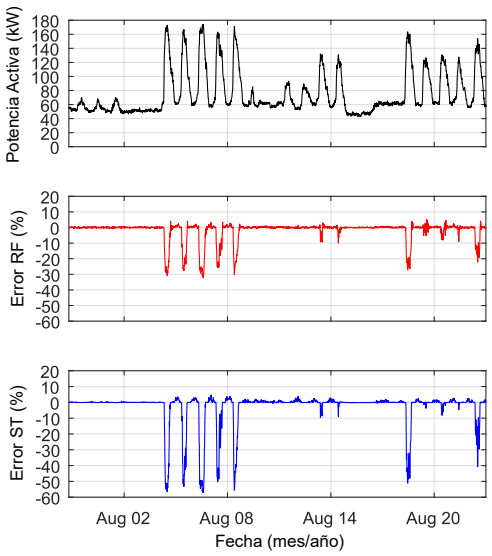

Figura 7: Base FCEyS, demanda primera quincena de agosto de 2025

tórico de la demanda, la presencia de eventos estructurales no reflejados en los datos de entrenamiento y la magnitud de la variabilidad propia en cada edificio.

5. Entrenamiento y resultados con aumento de datos

El aumento de datos se realizó en Python utilizando la biblioteca NumPy para la generación del ruido estocástico (Harris et al., 2020). Para ello se calculó la desviación estándar σ de la variable objetivo sobre el conjunto de entrenamiento y se generaron nuevas muestras mediante la adición de ruido gaussiano de amplitud controlada. Los modelos se entrenaron utilizando el conjunto aumentado, manteniendo los vectores de características sin modificación. Las perturbaciones se aplican únicamente a la variable objetivo del conjunto de entrenamiento, y no a las características derivadas de rezagos o ventanas deslizantes. Esto preserva la coherencia temporal de las características y evita inconsistencias en la estructura lag-target que podrían degradar el aprendizaje. El conjunto de test permanece completamente inalterado, garantizando que la evaluación del desempeño refleja la capacidad predictiva sobre datos reales.

La Fig. 8 compara los errores resultantes entre el pronóstico original y el obtenido con aumento de datos. No se observa una mejora respecto el escenario original, sino que en general el aumento de datos tiende a incrementar el error en intervalos donde originalmente era bajo. Más aún, este comportamiento también se puede apreciar en la Fig. 9, en la cual la magnitud del error era menor.

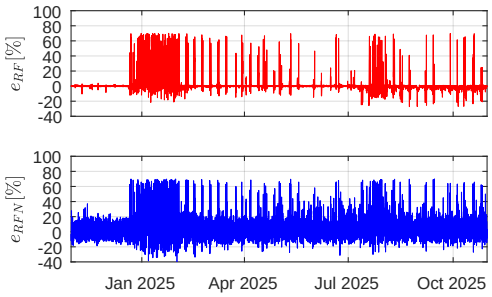

Figura 8: Base FI, error[%] original (rojo) vs aumentado (azul) con RF.

Por otro lado, la aplicación de la estrategia de aumento de datos al modelo Stacking produce mejoras en la calidad del pronóstico (Figuras 10 y 11), sobre todo respecto de las mayores desviaciones. En el caso de la Base FI, el piso del error aumenta levemente, aunque, la excursión máxima quedó por debajo de un 15 %. Con respecto a la Base FCEyS, los resultados son menos favorables, sin embargo, a costa de un incremento marginal en el error general, se logró reducir al 50 % la excursión máxima del error.

La Tabla 2 muestra las métricas de pronóstico (MAE, RMSE, MAPE, sMAPE y R^2) para cada edificio y modelo. Dado que la demanda neta de FI cae a valores cercanos a cero debido a la inyección fotovoltaica, afecta la interpretabilidad del MAPE debido a su sensibilidad a valores pequeños del denominador, lo

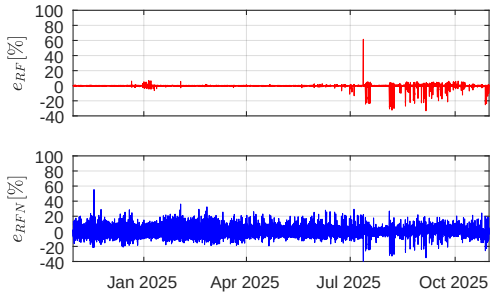

Figura 9: Base FCEyS, error[%] original (rojo) vs aumentado (azul) con RF.

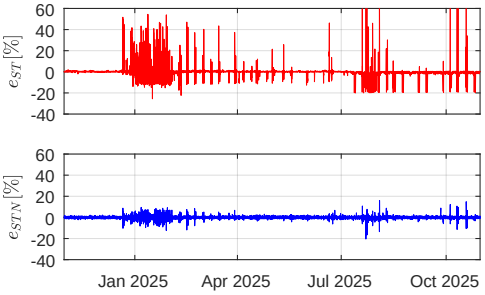

Figura 10: Base FI, error[%] original (rojo) vs aumentado (azul) con ST.

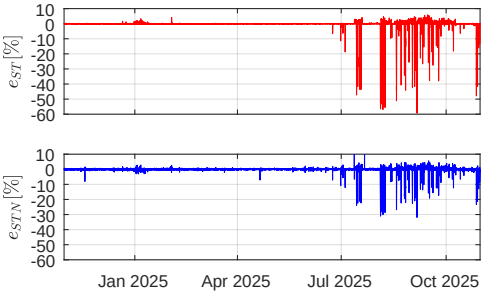

Figura 11: Base FCEyS, error[%] original (rojo) vs aumentado (azul) con ST.

que puede inducir errores relativos artificialmente elevados. Por este motivo, se priorizan el MAE y el RMSE como métricas de referencia y se reporta adicionalmente el sMAPE como una alternativa simétrica y y acotada, menos sensible a valores cercanos a cero (Hyndman y Koehler, 2006). Esto puede observarse en FI con RF, donde el MAPE (8.66 %) supera al sMAPE (7.22 %), diferencia que no aparece en FCEyS, donde ambos indicadores presentan valores muy similares. En FI, el modelo ST mejora al modelo RF en todas las métricas: el MAE se

reduce de 1.0054 a 0.1299 kW, el RMSE de 1.5228 a 0.2500 kW, el MAPE de 8.66 % a 0.61 %, y el coeficiente R^2 se aproxima a la unidad. Estos resultados confirman que ST corrige eficazmente el sesgo observado en RF bajo condiciones de baja demanda asociadas a la generación fotovoltaica. En FCEyS, ST también supera a RF, aunque con un comportamiento diferente. El MAE disminuye de 1.4748 a 0.4938 kW y los errores relativos (MAPE y sMAPE) se reducen de aproximadamente 2.45 % a 0.51 %. Sin embargo, la mejora observada en el RMSE es más moderada. Este contraste incrementa el cociente RMSE/MAE, lo que sugiere que, aunque ST reduce el error típico, persisten errores esporádicos de gran magnitud asociados a picos y transiciones abruptas de la demanda, los cuales tienen una influencia significativa sobre el término cuadrático del RMSE. En consecuencia, la mejora proporcionada por ST es generalizada sobre la mayor parte de la distribución de errores, aunque su impacto sobre eventos extremos resulta más limitado en FCEyS que en FI.

Tabla 2: Métricas de pronóstico por edificio y modelo.

Edificio	**Modelo**	**MAE** (kW)	**RMSE** (kW)	**MAPE** (%)	**sMAPE** (%)	$\mathbf{R^2}$
FI	RF	1.0054	1.5228	8.6568	7.2219	0.9865
FI	ST	0.1299	0.2500	0.6145	0.6192	0.9997
FCEyS	RF	1.4748	3.8811	2.4491	2.4629	0.9756
FCEyS	ST	0.4938	3.0024	0.5058	0.5138	0.9839

6. Discusión

Los resultados pueden interpretarse a partir de las propiedades de cada modelo. Random Forest reduce la varianza promediando árboles entrenados sobre submuestras. Al añadir ruido a la variable objetivo, el promedio absorbe gran parte de la perturbación, de modo que el aumento tiende a incrementar el sesgo sin una reducción compensatoria de la varianza, lo que explica la ausencia de mejoras, e incluso el deterioro, en su desempeño predictivo. En cambio, Stacking combina predicciones de boosting, más propensos a sobreajustar los mínimos históricos. El ruido aplicado sobre el objetivo actúa como una regularización que atenúa ese sobreajuste, y el meta-modelo lineal recalibra las salidas, reduciendo los errores extremos. Esto es consistente con que la mejora sea mayor en el perfil de mayor variabilidad (FCEyS, $\sigma = 21,1$ kW) que en FI ($\sigma = 14,8$ kW). Estos hallazgos son consistentes con la literatura: la degradación del pronóstico ante datos fuera del rango histórico refleja la dependencia de los métodos supervisados respecto de la representatividad del entrenamiento, y el aumento por ruido se enmarca en las revisiones técnicas para series temporales. En términos prácticos, los resultados sugieren que para la gestión energética universitaria conviene el reentrenamiento periódico ante cambios estructurales conocidos (ampliaciones fotovoltaicas, cambios de topología) y el uso de Stacking con aumento de datos cuando el error proviene de la variabilidad estacional de los recesos.

7. Conclusiones

En este artículo se evaluó el desempeño de los algoritmos Random Forest y Stacking para el pronóstico de demanda eléctrica en dos edificios universitarios con perfiles de consumo heterogéneos. Ambos modelos tienen una capacidad adecuada para reproducir la dinámica temporal de la demanda cuando las condiciones operativas permanecen dentro de los rangos históricamente observados; sin embargo, su precisión se ve afectada ante cambios estructurales, períodos de baja ocupación o desvíos significativos respecto del comportamiento registrado durante el entrenamiento. Se aplicó una técnica de aumento de datos sobre ambos métodos y ambos edificios. Los resultados evidencian un comportamiento dependiente del modelo: mientras que en Random Forest no se observan mejoras en el desempeño predictivo, en Stacking la técnica permite reducir de forma significativa las desviaciones extremas, a costa de un leve incremento del error promedio. El efecto del aumento de datos es, por lo tanto, específico del modelo, del edificio y del origen del error, resultando más efectivo cuando el error se asocia a la variabilidad estacional de los períodos de receso, mientras que su capacidad de corrección es limitada frente a errores derivados de cambios estructurales no representados en el entrenamiento. En conjunto, los resultados constituyen un paso más en el uso de aprendizaje automático en sistemas de gestión energética y ofrecen una base para el desarrollo de modelos más robustos y adaptativos en edificios públicos.

AGRADECIMIENTOS

Este trabajo fue soportado por la Universidad Nacional de Mar del Plata (UNMDP) y el CONICET. Los autores agradecen al Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) por el apoyo brindado a través de las Redes Temáticas RIPCEL 726RT0203 y RIIIAinSG 726RT0204.

Referencias

Breiman, L. (1996). Stacked Regressions. *Machine Learning*, *24*(1), 49-64. https://doi.org/10.1007/BF00117832

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. En C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 157-175). Springer. https://doi.org/10.1007/978-1-4419-9326-7_5

De Caro, F., Andreotti, A., Araneo, R., Panella, M., Rosato, A., Vaccaro, A., & Villacci, D. (2020). A Review of the Enabling Methodologies for Knowledge Discovery from Smart Grids Data. *Energies*, *13*(24). https://doi.org/10.3390/en13246579

Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, *74*, 902-924. https://doi.org/10.1016/j.rser.2017.02.085

Harris, C. R., et al. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

International Energy Agency [IEA]. (2025). Electricity Mid-Year Update 2025. https://www.iea.org/reports/electricity-mid-year-update-2025

Mahmood, M., Chowdhury, P., Yeassin, R., Hasan, M., Ahmad, T., & Chowdhury, N.-U.-R. (2024). Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. *Energy Conversion and Management: X*, *24*. https://doi.org/10.1016/j.ecmx.2024.100790

Maza Diaz, Y., Donato, P., Funes, M., & Orallo, C. (2024). Evaluating machine learning algorithms for power demand forecasting. *XV Latin-American Congress on Electricity Generation and Transmission (CLAGTEE 2024)*. https://clagtee.fi.mdp.edu.ar/full-papers-search-engine/papers/ID061.pdf

Maza Diaz, Y., Donato, P., Funes, M., & Orallo, C. (2025). Evaluación de algoritmo Random Forest para predicción de demanda en suministros de energía de la UNMDP. *VI JORNADAS DE INGENIERíA APLICADA*. https://owncloud.fi.mdp.edu.ar/index.php/s/yE7RyNxy1kNg8w2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Powell, J., McCafferty-Leroux, A., Hilal, W., & Gadsden, S. A. (2024). Smart grids: A comprehensive survey of challenges, industry applications, and future trends. *Energy Reports*, *11*, 5760-5785. https://doi.org/10.1016/j.egyr.2024.05.051

Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters*, *24*(3), 279-283. https://doi.org/10.1109/lsp.2017.2657381

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2021). Time Series Data Augmentation for Deep Learning: A Survey. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 4653-4660. https://doi.org/10.24963/ijcai.2021/631