

Pronóstico del Consumo de Energía Eléctrica en Edificio Inteligente



Colaboración

Yianko de Jesús Castellanos Pol Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad Académica Tepic; Ansel Yoan Rodríguez González; Yoan Martínez López, Universidad de Córdoba

Fecha de recepción: 17 de septiembre 2025

Fecha de aceptación: 17 de septiembre de 2025

RESUMEN: La optimización del consumo energético en edificios inteligentes es esencial para reducir el impacto ambiental y mejorar la eficiencia en el uso de recursos. En este trabajo se aborda la predicción del consumo eléctrico a partir de datos históricos obtenidos de la Competition on Electric Energy Consumption Forecast 2025. El estudio se organizó en cuatro etapas: recopilación de datos, preprocesamiento, normalización y experimentación con diferentes modelos de aprendizaje automático. Se evaluaron cuatro algoritmos representativos: Regresión Lineal, K-Nearest Neighbors (KNN), Random Forest y Máquinas de Soporte Vectorial (SVM). El rendimiento se midió utilizando métricas estándar como MSE, RMSE, MAE y R2. Los resultados muestran que Random Forest alcanzó el mejor desempeño, con un $R^2=0.88$ y errores mínimos tras la normalización de los datos. Estos hallazgos confirman la efectividad de los métodos de ensamble para capturar patrones complejos de consumo y refuerzan su potencial como herramienta robusta para la gestión energética en tiempo real en edificios inteligentes.

PALABRAS CLAVE: predicción, energía, consumo, edificios inteligentes, aprendizaje automático.

ABSTRACT: The optimization of energy consumption in smart buildings is essential to reduce environmental impact and improve resource efficiency. This work addresses the prediction of electricity consumption using historical data from the Competition on Electric Energy Consumption Forecast 2025. The study was structured in four stages: data collection, preprocessing, normalization, and experimentation with different machine learning models. Four representative algorithms were evaluated: Linear Regression, K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machines (SVM). Performance was assessed using standard metrics such as MSE, RMSE, MAE, and R2. The results show that Random Forest achieved the best performance, with an $R^2=0.88$ and minimal errors after data normalization. These findings confirm the effectiveness of ensemble methods in capturing complex consumption patterns and highlight their potential as robust tools for real-time energy management in smart buildings.

KEYWORDS: forecasting, energy, consumption, smart buildings, machine learning.

INTRODUCCIÓN

En los últimos años, la optimización del consumo energético en edificios inteligentes ha ganado relevancia, motivada por la necesidad de reducir el impacto ambiental y mejorar la eficiencia energética [1]. Esta importancia se refleja incluso en la creación de competencias internacionales, como la Competition on Electric Energy Consumption Forecast 2025 , que buscan promover el desarrollo de modelos precisos de predicción del consumo energético. Para ello, se han desarrollado modelos de forecasting como ARIMA, LSTM, redes neuronales profundas y modelos híbridos [2]. La pre-

dicción de la demanda energética resulta esencial para gestionar eficientemente los recursos, especialmente en sistemas que integran energías renovables, como la fotovoltaica [3], [4].

En este trabajo se exploran modelos clásicos como Regresión Lineal [5], K-Nearest Neighbors (KNN) [6], Random Forest [7] y Máquinas de Soporte Vectorial (SVM) [8], debido a su interpretabilidad, menor complejidad computacional y capacidad de ofrecer resultados robustos con cantidades moderadas de datos históricos. Estos modelos permiten implementar soluciones efectivas y comprensibles para la predicción del consumo energético, sin requerir la infraestructura y los grandes volúmenes de datos que exigen las técnicas más complejas.

La investigación se estructura en diferentes fases que abarcan desde la recopilación y preparación de los datos hasta la evaluación experimental de los modelos seleccionados, con el propósito de determinar cuál ofrece un desempeño más adecuado en el contexto del edificio analizado.

MATERIAL Y MÉTODOS

El desarrollo del presente trabajo se organiza en cuatro etapas fundamentales: identificación del consumo energético y sus principales características, preprocesamiento de los datos, recopilación y análisis de los mismos y experimentación con diferentes modelos de predicción.

Conjunto de Datos

Los datos utilizados fueron obtenidos de la Competition on Electric Energy Consumption Forecast 2025. Esta base de datos se encuentra distribuida en cinco archivos CSV: uno con un año de registro, otro con 40 días y tres conjuntos correspondientes a tres días de consumo energético. Los archivos incluyen información de la temperatura exterior, condiciones climáticas y la energía generada por los paneles fotovoltaicos del edificio.

Cabe señalar que la documentación de la Competition on Electric Energy Consumption Forecast 2025 no incluye una descripción detallada del edificio, solo que se trata de un smart building con paneles fotovoltaicos y sensores ambientales. En consecuencia, el presente análisis se realiza considerando únicamente la información disponible en la base de datos pública y reconociendo esta limitación.

La Tabla 1 presenta los promedios de consumo, temperatura y generación fotovoltaica en los diferentes períodos analizados. Se observa que la demanda del edificio varía entre 1,560.17 y 2,076.67 kWh, mientras que la generación proveniente de los paneles fotovoltaicos fluctúa entre 236.19 y 714.74 kWh, cubriendo aproximadamente el 19% de la alimentación total del edificio. Los valores de temperatura se mantienen relativamente constantes.

Tabla 1. Promedio de consumos, temperatura y generación en diferentes períodos.

| Ficheros | Consumos (kWh) | Temperatura (°C) | Generación (kW) |
|----------|----------------|------------------|-----------------|
| Año | 1909.00 | 16.97 | 714.74 |
| 40 días | 2076.67 | 12.51 | 257.76 |
| Día 1 | 1656.38 | 13.74 | 240.35 |
| Día 2 | 1560.17 | 14.62 | 238.24 |
| Día 3 | 1722.67 | 15.23 | 236.19 |

Fuente: Elaboración propia.

Las Figuras 1 y 2 muestran las tendencias generales del consumo energético anual y de 40 días, evidenciando variaciones crecientes o decrecientes que reflejan cambios en la demanda a lo largo del tiempo. Estas fluctuaciones pueden estar asociadas a eventos estacionales o extraordinarios que alteran el patrón habitual de consumo.

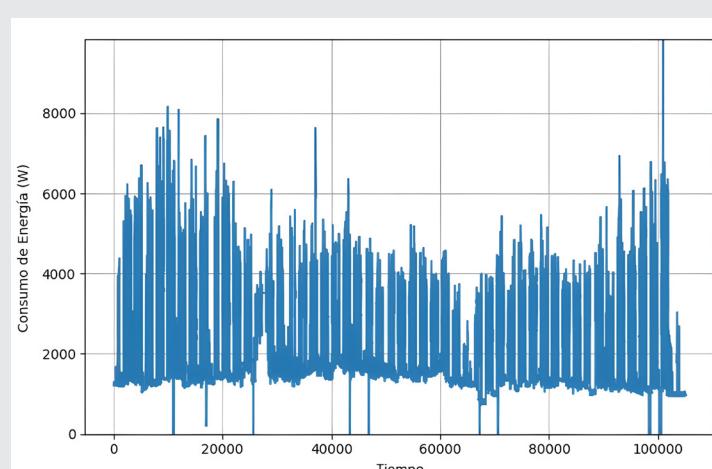


Figura 1. Consumo de Energía por día en un año.

Fuente: Elaboración propia.

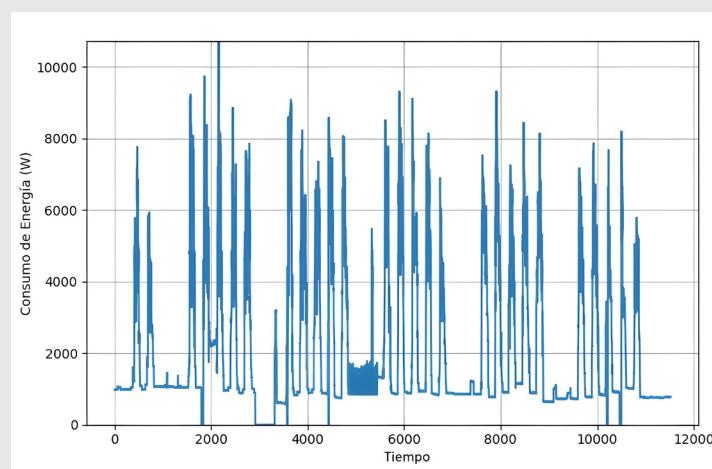


Figura 2. Consumo de Energía por día en 40 días.

Fuente: Elaboración propia.

La Figura 3 presenta los consumos diarios para los días 1, 2 y 3, mostrando una variación relativamente baja entre días consecutivos.

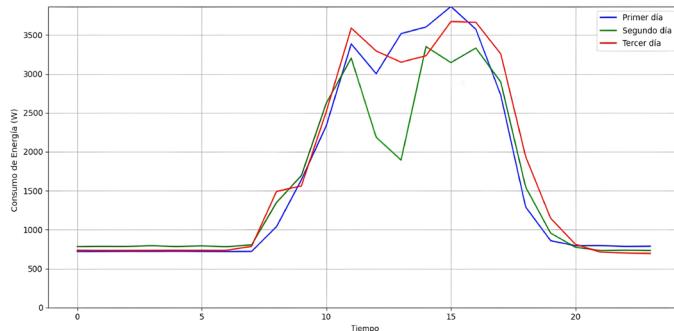


Figura 3. Consumo de Energía por días 1, 2 y 3.

Fuente: Elaboración propia.

Preprocesamiento de los Datos

El preprocesamiento consistió en reducir los datos del período anual, pasando de un muestreo cada 5 minutos a intervalos de 1 hora, calculando el consumo promedio por hora. Esta transformación permite simplificar el análisis y facilita la visualización e interpretación de patrones de consumo a diferentes escalas temporales.

Como paso opcional dentro del preprocesamiento, se aplicó normalización de las variables, con el objetivo de homogeneizar las escalas y mejorar la convergencia de algunos modelos de predicción sensibles a la magnitud de los datos, como KNN y SVM [9].

De esta forma, se obtuvieron dos conjuntos de datos:

- Conjunto original preprocesado: datos reducidos a intervalos de 1 hora sin normalización
- Conjunto normalizado: datos preprocesados con normalización aplicada a todas las variables.

En los experimentos, ambos conjuntos fueron utilizados.

Recopilación y Análisis de Datos

La experimentación se centró en evaluar el rendimiento de los modelos de predicción del consumo energético:

- Regresión Lineal [5]: modelo paramétrico que establece una relación lineal entre las variables predictivas y la variable objetivo mediante mínimos cuadrados.
- K-Nearest Neighbors (KNN) [6]: modelo no paramétrico que predice el valor de una instancia en función de los k vecinos más cercanos en el conjunto de entrenamiento.
- Random Forest [7]: ensamble de árboles de decisión que mejora la precisión y reduce el sobreajuste mediante votación promedio de múltiples árboles.

- Máquinas de Soporte Vectorial (SVM) [8]: algoritmo supervisado que encuentra un hiperplano óptimo para separar o predecir los datos, maximizando el margen y utilizando funciones kernel para relaciones no lineales.

Se utilizaron las dos versiones del conjunto de datos: Conjunto original preprocesado y Conjunto normalizado. En cada modelo se emplearon procedimientos de ajuste y validación mediante división de los datos en conjuntos de entrenamiento y prueba (80%-20%), con el objetivo de evaluar su capacidad de generalización. Los parámetros principales como el número de vecinos en KNN, la cantidad de árboles en Random Forest y los hiperparámetros C y γ en SVM, fueron determinados empíricamente a partir de pruebas iterativas para optimizar el rendimiento. Para medir la precisión de los modelos se emplearon las siguientes métricas [10]:

Raíz del Error Cuadrático Medio ($RMSE$)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_m - P_c)^2}$$

Error Cuadrático Medio (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^N (P_m - P_c)^2$$

Error Absoluto Medio (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_m - P_c|$$

Coeficiente de Determinación (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_m - P_c)^2}{\sum_{i=1}^N (P_m - \bar{P}_c)^2}$$

RESULTADOS

La Tabla 2 muestra la comparación de las métricas de desempeño de los modelos evaluados, tanto para los datos originales como para los normalizados. Se observa que la normalización mejora significativamente los resultados en todos los casos, particularmente en KNN y Random Forest.

Regresión Lineal

Los resultados obtenidos mediante Regresión Lineal reflejan un rendimiento deficiente: el modelo presentó un $R^2=-0.16$, lo que sugiere que no logra explicar la variabilidad de los datos. Además, las métricas de error fueron elevadas ($MSE = 1.13$, $RMSE = 1.16$ y $MAE = 0.98$), indicando que las predicciones se alejan considerablemente de los valores reales. Estos resultados ponen de manifiesto que, incluso tras la normalización, la Regresión Lineal no es adecuada para este problema de predicción (Tabla 2).

Tabla 2. Comparación de métricas entre métodos para conjunto original preprocesado y conjunto normalizado.

| Método | Métrica | Conjunto original preprocesado | Conjunto normalizado |
|-------------------------------|----------------|--------------------------------|----------------------|
| Regresión Lineal | MSE | 637475.27 | 0.28 |
| | RMSE | 858.18 | 0.53 |
| | MAE | 689.81 | 0.40 |
| | R ² | 0.48 | 0.71 |
| K-Nearest Neighbors | MSE | 140373.31 | 0.12 |
| | RMSE | 374.66 | 0.35 |
| | MAE | 264.68 | 0.25 |
| | R ² | 0.90 | 0.88 |
| Random Forest | MSE | 1603928.13 | 1.13 |
| | RMSE | 1266.46 | 1.16 |
| | MAE | 1111.10 | 0.98 |
| | R ² | -0.13 | -0.16 |
| Máquinas de Soporte Vectorial | MSE | 2508191.02 | 1.70 |
| | RMSE | 1583.73 | 1.30 |
| | MAE | 1168.19 | 1.04 |
| | R ² | -0.77 | -0.74 |

Fuente: Elaboración propia.

K-Nearest Neighbors (KNN)

El modelo KNN se ajustó utilizando un valor de $k=10$. Los resultados muestran un desempeño moderado, con un coeficiente de determinación $R^2=0.71$, lo que indica una capacidad limitada para explicar la variabilidad de los datos. Las métricas de error ($MSE = 0.28$, $RMSE = 0.53$ y $MAE = 0.40$) son relativamente altas, lo que refleja predicciones con errores considerables. Tanto el RMSE como el MAE confirman que el modelo presenta una precisión restringida. En general, el ajuste de parámetros permitió mejorar los resultados respecto al conjunto sin normalización, aunque el rendimiento sigue siendo modesto (Tabla 2).

Random Forest

El modelo se entrenó con 100 árboles de decisión. Los resultados (Tabla 2) muestran un desempeño sobresaliente, con un $R^2=0.88$, lo que evidencia una alta capacidad explicativa de la variabilidad de los datos. Las métricas de error fueron bajas ($MSE = 0.12$, $RMSE = 0.35$ y $MAE = 0.25$), confirmando que el modelo produce predicciones consistentes y precisas. La normalización de los datos contribuyó a reducir aún más los errores, consolidando a Random Forest como el modelo más robusto entre los evaluados.

Support Vector Machine (SVM)

En este trabajo SVM se configuró con $C=100$, $\gamma=scale$ y $\epsilon=0.1$. A pesar de ello, el modelo presentó un desempeño muy pobre, con $R^2=-0.74$, lo que evidencia que no logra capturar la variabilidad de los datos. Asimismo, las métricas de error fueron elevadas ($MSE = 1.70$, $RMSE = 1.30$ y

$MAE = 1.04$), lo que confirma que las predicciones difieren significativamente de los valores reales.

Gráfica de Ajuste

La Figura 4 muestra la relación entre los valores reales de consumo energético (kWh) y los valores predichos por el método Random Forest. Se observa un alto grado de correspondencia entre ambas series, lo que indica que el modelo logró capturar de manera adecuada las tendencias y patrones presentes en los datos.

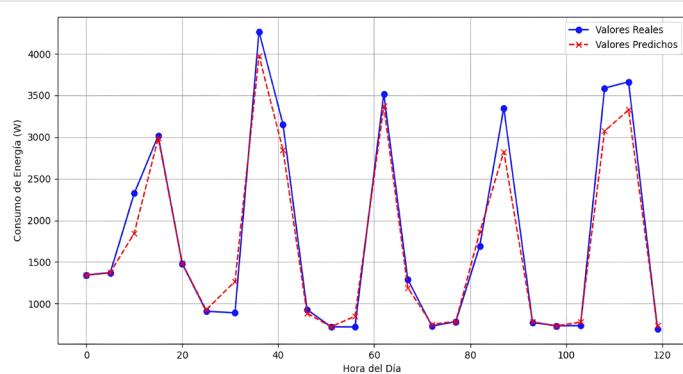


Figura 4. Comparación entre consumos reales y predichos mediante Random Forest.

Fuente: Elaboración propia.

En particular, Random Forest, al estar basado en múltiples árboles de decisión, demuestra una capacidad superior para modelar relaciones complejas entre las variables. Esta característica lo convierte en una herramienta adecuada para la optimización de la gestión del consumo energético en tiempo real y para apoyar la toma de decisiones estratégicas.

Discusión de resultados

Los resultados obtenidos evidencian diferencias notables en el desempeño de los modelos evaluados. En particular, Random Forest presentó un ajuste significativamente superior frente a los demás algoritmos, alcanzando un $R^2=0.88$ y un $MSE = 0.12$ en el conjunto normalizado, lo que confirma su capacidad para capturar la variabilidad del consumo energético en el edificio analizado. Estos resultados refuerzan la idoneidad de los métodos basados en ensambles en problemas de predicción de series temporales con múltiples variables y relaciones no lineales. Asimismo, la normalización de los datos resultó ser un factor determinante, mejorando el rendimiento de modelos sensibles a la escala, como KNN y SVM, y optimizando también los resultados de Random Forest. En contraste, la Regresión Lineal y las Máquinas de Soporte Vectorial (SVM) mostraron un desempeño deficiente, incluso después de aplicar normalización, lo que pone de manifiesto sus limitaciones para capturar patrones de consumo caracterizados por alta variabilidad y no linealidad. Finalmente, cabe señalar que, si bien no se evaluaron técnicas de aprendizaje profundo como LSTM

o CNN-LSTM, los resultados alcanzados sugieren que los modelos clásicos bien configurados ofrecen un balance favorable entre precisión, interpretabilidad y eficiencia computacional, especialmente en escenarios con bases de datos de tamaño intermedio.

En términos prácticos, la comparación de algoritmos realizada en este estudio aporta una referencia útil para investigadores y profesionales que buscan seleccionar técnicas de aprendizaje automático adecuadas a distintos escenarios de consumo energético, favoreciendo el desarrollo de soluciones escalables y sostenibles para la gestión inteligente de la energía.

CONCLUSIONES

En este trabajo se demostró que Random Forest constituye el modelo más efectivo para la predicción del consumo energético en edificios inteligentes, superando en rendimiento a KNN, Regresión Lineal y SVM, gracias a su capacidad para modelar relaciones complejas y no lineales entre las variables. Los experimentos confirmaron que la normalización de datos es un paso fundamental en el preprocesamiento, pues mejora el desempeño de los algoritmos sensibles a la escala y contribuye a reducir los errores de predicción. Por el contrario, los modelos lineales y de márgenes presentaron un rendimiento limitado, al no capturar adecuadamente la complejidad de los patrones de consumo.

En conjunto, los resultados obtenidos consolidan la utilidad de los métodos basados en ensambles como herramientas robustas, precisas y eficientes para la gestión del consumo energético en tiempo real, y plantean la necesidad de explorar en futuros trabajos la integración de técnicas de aprendizaje profundo e híbrido que combinen precisión, escalabilidad y aplicabilidad en entornos reales de edificios inteligentes. Además, los resultados obtenidos proporcionan una base comparativa útil para investigadores interesados en el modelado energético, facilitando la validación de nuevos enfoques y la reproducción de estudios en distintos contextos de edificios inteligentes.

BIBLIOGRAFÍA

- [1] H. P. Das, Y.-W. Lin, U. Agwan, L. Spangher, A. Devonport, Y. Yang, J. Drgona, S. Schiavon, and C. J. Spanos, “Machine Learning for Smart and Energy-Efficient Buildings,” *arXiv preprint arXiv:2211.14889*, 2022.
- [2] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, “Energy forecasting: A review and outlook,” *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.
- [3] H. Yin, M. O'Neill, M. Hossain, and Y. Liu, “A comprehensive review of neural networks for building energy consumption forecasting,” *Sustainability*, vol. 16, no. 17, p. 7805, 2024.
- [4] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, “A review of deep learning for renewable energy forecasting,” *Energy Conversion and Management*, vol. 198, p. 111799, Oct. 2019.
- [5] H. Yang, M. Ran, and C. Zhuang, “Prediction of building electricity consumption based on Joinpoint-multiple linear regression,” *Energies*, vol. 15, no. 22, p. 8543, 2022.
- [6] G. Hong, G.-S. Choi, J.-Y. Eum, H. S. Lee, and D. D. Kim, “The hourly energy consumption prediction by KNN for buildings in community buildings,” *Buildings*, vol. 12, no. 10, p. 1636, 2022,
- [7] C. B. Pop, V. R. Chifu, C. Cordea, E. S. Chifu, and O. Barsan, “Forecasting the short-term energy consumption using random forests and gradient boosting,” *Energy Reports*, vol. 8, pp. 1–10, 2022.,
- [8] Z. Meng, X. Zhang, and Y. Wang, “Forecasting energy consumption based on support vector regression and Markov chain,” *Frontiers in Environmental Science*, vol. 10, p. 883711, 2022.,
- [9] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pinetelas, “Data preprocessing for supervised learning,” *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [10] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature,” *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014

