

Previsión del consumo eléctrico en el cantón Salcedo mediante técnicas de aprendizaje automático

Electricity consumption forecasting in the canton of Salcedo using machine learning techniques

Fecha de recepción: 2023-11-20 • Fecha de aceptación: 2024-01-17 • Fecha de publicación: 2024-02-10

Oscar Fabricio Chicaiza Yugcha¹

Instituto Superior Tecnológico Simón Bolívar, Ecuador

o_chicaiza@istsb.edu.ec

<https://orcid.org/0000-0002-4170-2186>

Carlos Julio Martínez Guaman²

Instituto Superior Tecnológico Simón Bolívar, Ecuador

c_martinez@istsb.edu.ec

<https://orcid.org/0000-0001-6832-902X>

Israel Antonio Orozco Manobanda³

Instituto Superior Tecnológico Simón Bolívar, Ecuador

i_orozco@istsb.edu.ec

<https://orcid.org/0000-0003-2931-0240>

Ángel Danilo Arellano Castro⁴

Instituto Superior Tecnológico Simón Bolívar, Ecuador

a_arellano@istsb.edu.ec

<https://orcid.org/0000-0002-0966-7122>

RESUMEN

En respuesta al crecimiento de la demanda de energía eléctrica, este estudio se centra en la eficiente previsión del consumo eléctrico en el cantón Salcedo, Ecuador. Se adoptaron las técnicas de aprendizaje automático Random Forest y XGBoost para pronosticar la demanda de seis parroquias en el sector residencial con registros de enero de 2017 hasta diciembre de 2022. La metodología abarcó la recopilación de datos, preprocesamiento, entrenamiento y evaluación de modelos. Se utilizaron métricas como el RMSE y MAPE para validar el rendimiento, destacando Random Forest como el más eficaz en la previsión de la demanda en todas las parroquias, mostrando una adaptación más adecuada a las peculiaridades del consumo eléctrico. Este enfoque no solo proporciona una base para la planificación eficiente de la generación y distribución de energía, sino que también destaca la utilidad de las técnicas de aprendizaje automático en entornos de previsión de consumo energético.

PALABRAS CLAVE: previsión de consumo eléctrico, aprendizaje automático, Random Forest, XGBoost, Ecuador

ABSTRACT

In response to the growth of electricity demand, this study focuses on the efficient forecasting of electricity consumption in Salcedo canton, Ecuador. Random Forest and XGBoost machine learning techniques were adopted to forecast the demand of six parishes in the residential sector with records from January 2017 to December 2022. The methodology encompassed data collection, preprocessing, training, and model evaluation. Metrics such as RMSE and MAPE were used to validate performance, highlighting Random Forest as the most effective in forecasting demand in all parishes, showing a more adequate adaptation to the peculiarities of electricity consumption. This approach not only provides a basis for efficient power generation and distribution planning, but also highlights the usefulness of machine learning techniques in energy consumption forecasting environments.

KEYWORDS: power forecasting, machine learning, Random Forest, XGBoost, Ecuador

Introducción

El desarrollo a nivel mundial ha contribuido significativamente al aumento de la demanda de energía eléctrica, impulsado tanto por el crecimiento demográfico, como por la proliferación de dispositivos inteligentes y la adopción de vehículos eléctricos (Bedi & Toshniwal, 2019). Por esta razón, contar con una proyección precisa de la demanda de electricidad desempeña un papel crucial en la gestión y generación de los sistemas eléctricos. Al anticipar la demanda máxima de energía eléctrica, las compañías de servicios públicos pueden llevar a cabo una planificación más eficiente de la generación y distribución de energía (R. Li et al., 2020). Es esencial que las empresas de generación integren técnicas de previsión de carga para proyectar de manera efectiva la futura demanda de electricidad (Mir et al., 2020).

Dentro del ámbito de la previsión del consumo energético se han empleado enfoques clásicos basados en modelos econométricos y estadísticos. Sin embargo, se ha constatado que el aprendizaje automático o redes neuronales puede superar a las metodologías tradicionales en escenarios que involucran series temporales impredecibles o inusuales (Mañay et al., 2022; Shin & Woo, 2022).

La previsión del consumo de energía implica la utilización de diversas estrategias. Investigaciones posteriores han examinado el uso de técnicas de aprendizaje automático para predecir el consumo de energía. Los modelos de previsión de carga basados en aprendizaje automático han ganado popularidad y utilizan datos históricos para prever la demanda futura (Ruiz-Abellón et al., 2018).

1.1 Modelos de aprendizaje automático supervisado

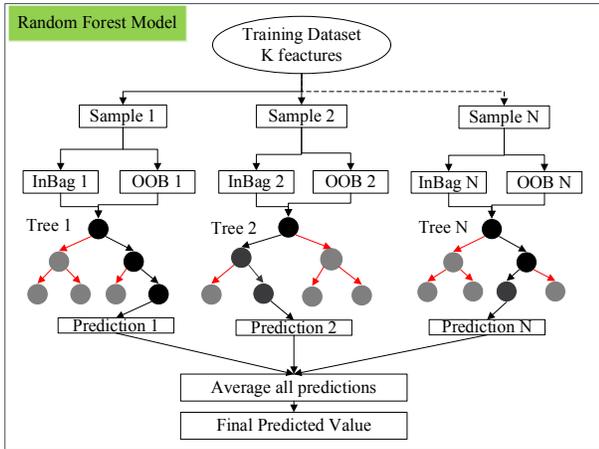
El aprendizaje automático se centra en el reconocimiento de patrones a través del entrenamiento con datos históricos para predecir o identificar patrones. El rendimiento tiende a mejorar con la disponibilidad de un conjunto de observaciones lo suficientemente amplio, dependiendo de la aplicación en desarrollo (Rojas, 2020). Esta investigación se enfoca en los modelos Random Forest y XGBoost para la previsión del consumo de energía.

1.1.1 Random Forest (RF)

RF es un modelo de predicción de conjunto que consta de una colección de diferentes árboles de regresión que se entrenan mediante bagging y selección aleatoria de variables (Ahmad et al., 2018). En la *Figura 1* se presenta el modelo de Random Forest de regresión.

Figura 1

Estructura del Modelo Random Forest Regresor



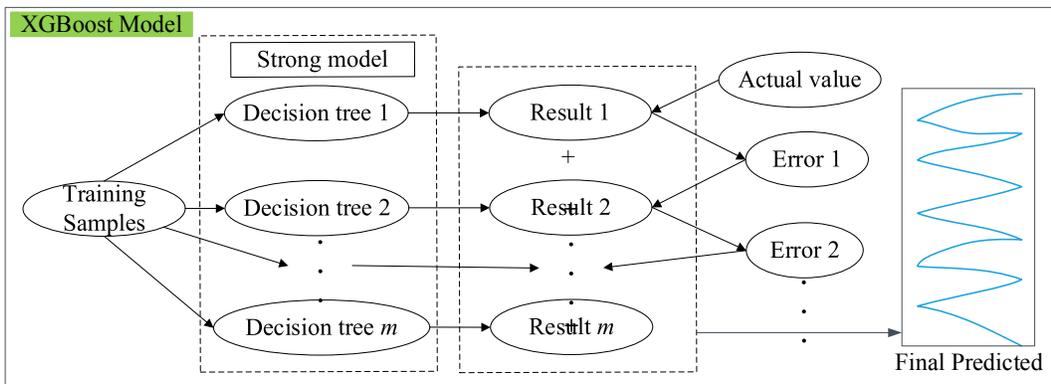
1.1.2 Extreme Gradient Boosting (XGBoost)

El Extreme Gradient Boosting (Refuerzo de Gradientes Extremo) es una técnica reconocida por su capacidad para lograr alta precisión y eficacia al trabajar con datos organizados. Este enfoque opera mediante la técnica de aumento de gradiente, la cual combina árboles de decisión para crear un modelo de conjunto potente (Dhaliwal et al., 2018).

En lugar de crear varios modelos por separado como en el Random Forest, XGBoost construye un solo modelo, pero lo hace en etapas iterativas. En cada iteración se ajusta un nuevo modelo para corregir los errores cometidos por los modelos anteriores (X. Li et al., 2022). En la *Figura 2* se muestra el modelo XGBoost.

Figura 2

Estructura del Modelo XGBoost Regresor



1.1.3 Métricas de evaluación

La tarea clave en esta investigación es la elección de un modelo que tenga la capacidad óptima para predecir resultados en nuevos conjuntos de datos. En este contexto se emplean **métricas ampliamente aceptadas** para la evaluación de los modelos de regresión. Dichas métricas engloban el error de la raíz cuadrada de la media (RMSE), que evalúa la variabilidad entre valores reales y pronosticados, y el error porcentual absoluto medio (MAPE), que cuantifica el error porcentual entre los valores reales y predichos del modelo (Banik et al., 2021). En las Ecuaciones (1) y (2) del RMSE y MAPE se presentan en las Ecuaciones (1) y (2), respectivamente.

Error de la raíz cuadrada de la media (RMSE):

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

Error porcentual absoluto de la media (MAPE):

$$MAPE = \left(\frac{\sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}}{n} \right) * 100 \quad (2)$$

Donde n es el número de datos, y_i es la carga actual en el instante i , y \hat{y}_i es la carga pronosticada en el instante i .

1.2 Problemática

Hoy en día, la planificación eficiente para la generación y transmisión de energía es crucial para evitar apagones o racionamientos, situaciones que lamentablemente están ocurriendo actualmente en el país. Contar con mecanismos de previsión de energía mediante técnicas de aprendizaje automático se presenta como una alternativa valiosa. La disponibilidad de datos históricos de consumo permite establecer una estructura sólida para la producción y transmisión de energía, respaldando así la implementación de técnicas avanzadas de aprendizaje automático. Entre estas técnicas se destacan algoritmos como Random Forest y XGBoost, que aprovechan la diversidad de múltiples modelos para mejorar significativamente la precisión y robustez de las predicciones. Random Forest se centra en la creación de un conjunto independiente de árboles y promedia sus predicciones, mientras que XGBoost emplea boosting secuencial para ajustar un solo modelo a lo largo de múltiples iteraciones y corregir errores de manera eficaz.

En el contexto de la previsión de la demanda de energía eléctrica en el cantón Salcedo, Ecuador se utilizan datos recopilados desde enero de 2017 hasta diciembre de 2022. Este estudio se enfoca en los clientes residenciales de las seis parroquias del cantón: San Miguel, Antonio José Holguín (Santa Lucía), Cusubamba, Mulalillo, Mulliquindil (Santa Ana), y Panzaleo. El objetivo principal es evaluar la efectividad de las técnicas Random Forest (RF) y XGBoost en la previsión del consumo de energía a lo largo de una proyección de un año.

Las siguientes secciones de la investigación están organizadas de la siguiente manera. En la sección II se describen los datos empleados y el método aplicado para la generación y comparación de modelos de pronóstico. Luego, en la sección III se presentan resultados y comparaciones entre los modelos elaborados. Finalmente, en la sección IV se exponen las respectivas conclusiones.

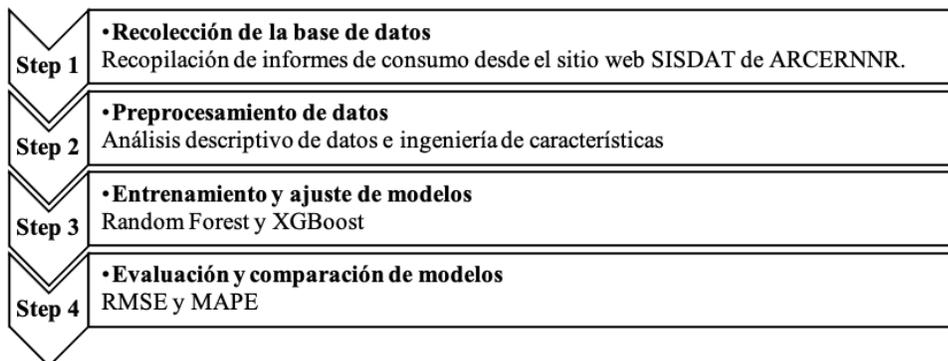
Metodología

En esta sección se proporciona una descripción detallada del enfoque metodológico empleado en la previsión del consumo de energía en el cantón Salcedo. La estrategia adoptada sigue la metodología de la Estructura de Descomposición del Trabajo (WBS, por sus siglas en inglés), que facilita la división de proyectos en componentes individuales (Mañay Chochos et al., 2022).

La metodología se divide en diversas etapas: recopilación de información desde bases de datos, preprocesamiento de datos, entrenamiento de modelos, ajuste de modelos y evaluación. La *Figura 3* ilustra el procedimiento del estudio, presentando los pasos a seguir en el proceso de previsión del consumo energético.

Figura 3

Metodología Propuesta para la Previsión de Consumo Energético

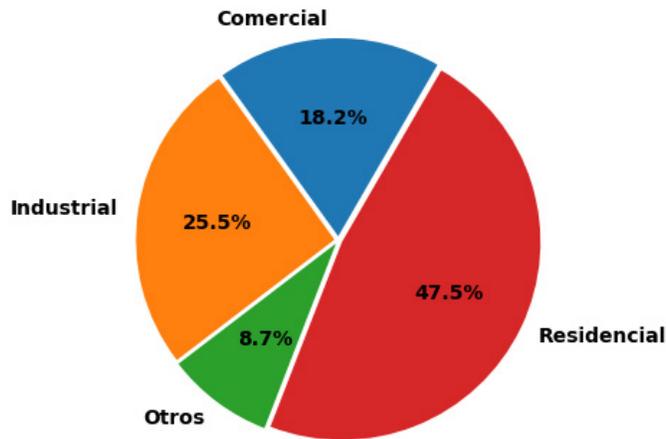


2.1 Recolección de la base de datos

En este estudio se utilizan informes mensuales que contienen datos de clientes regulados sobre el consumo energético en las parroquias del cantón Salcedo. El conjunto de informes abarca desde enero de 2017 hasta diciembre de 2022. En la *Figura 4* se presentan los porcentajes de consumo correspondientes a cada tipo de cliente en el cantón, destacando que el cliente residencial registra el mayor consumo, alcanzando el 47.5% en las parroquias. Estos informes están disponibles en el portal web de la Sistematización de Datos del Sector Eléctrico (SISDAT) de la Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables (ARCERNNR), una entidad gubernamental ecuatoriana (ARCERNNR, 2023).

Figura 4

Segmentación por Porcentajes del Consumo en el Cantón



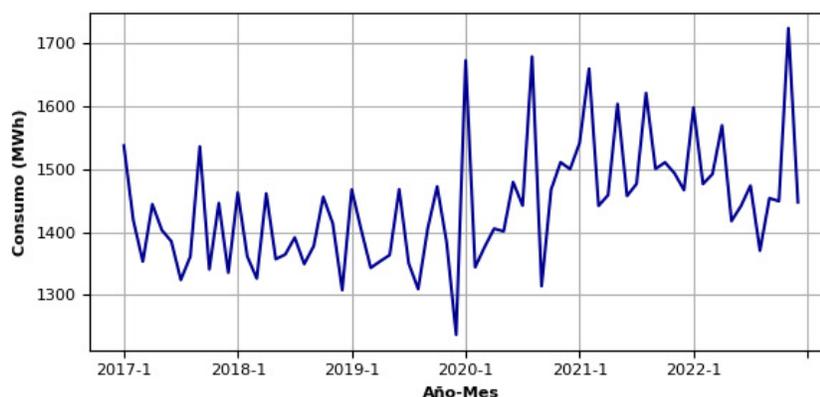
2.2 Preprocesamiento de datos

Se creó una matriz multidimensional que cuenta con un total de $n = 432$ muestras y $p = 6$ características, que abarcan aspectos como año, mes, parroquia, número de clientes, energía consumida y facturación (Carrillo et al., 2023).

Se llevó a cabo la depuración de registros que contenían valores iguales o inferiores a cero. Respecto a las variables categóricas, se aplicó la técnica de etiquetado numérico a la columna Mes (1 a 12), y para Parroquia se empleó la estrategia de codificación One Hot Encoding para lograr la estandarización necesaria (Guachimboza-Davalos et al., 2021; Rosero-Garcia et al., 2021). Después de completar la depuración se mantuvo el tamaño original de la matriz. La *Figura 4* presenta la tendencia de consumo a lo largo del período de estudio de los clientes residenciales en el cantón Salcedo.

Figura 5

Tendencia del Consumo Eléctrico Período: 2017-1 a 2022-12



2.3 Modelos de pronóstico: entrenamiento y evaluación

En esta fase se proporcionarán detalles sobre los segmentos empleados, tanto en el entrenamiento como en la evaluación de los modelos. La implementación de dichos modelos se llevó a cabo utilizando Python junto con la biblioteca Scikit-learn. Es importante señalar que los registros mensuales abarcan un periodo de seis años, desde enero de 2017 hasta diciembre de 2023. El enfoque central de la investigación se centra en la predicción del consumo eléctrico en el cantón Salcedo, específicamente en sus seis parroquias.

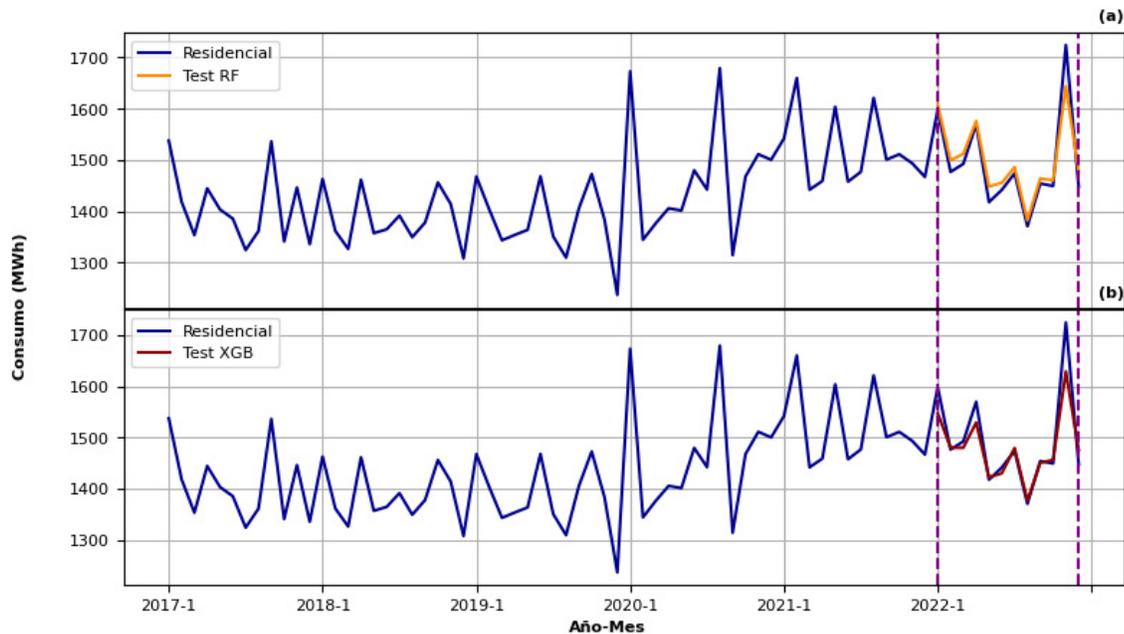
En la etapa de entrenamiento se seleccionaron datos comprendidos entre enero de 2017 y diciembre de 2021, reservando los datos correspondientes al año 2022 para la fase de prueba. Cinco predictores, a saber, año, mes, parroquia, número de clientes y facturación fueron elegidos como características independientes para el entrenamiento de los modelos. En este contexto, la columna de Energía Consumida se designa como la variable dependiente.

Para estructurar el modelo de entrenamiento tanto para Random Forest como para XGBoost, es necesario invocar inicialmente a las clases RandomForestRegressor y XGBRegressor, respectivamente. Durante este proceso se ajustan hiperparámetros cruciales, como el número de estimadores (`n_estimators`). Además, para mejorar la reproducibilidad de los resultados y garantizar un ajuste más estable del algoritmo, se establece el parámetro `random_state` en 42.

Con el propósito de identificar el modelo que mejor se ajuste y logre el RMSE y MAPE más bajos, se emplea una transformación utilizando el escalador MinMaxScaler para ajustar las observaciones en un rango de $[-1, +1]$ para Random Forest y $[0, +1]$ para XGBoost, con el fin de mejorar el desempeño del modelo. Se observa que se alcanza el mejor rendimiento en la reducción de MAPE y RMSE mediante un árbol que consta de 500 estimadores. La *Figura 6* presenta una comparación gráfica entre los valores reales y las predicciones del consumo de energía para ilustrar estos resultados.

Figura 6

Modelos de Pronóstico con Random Forest y XGBoost



Resultados

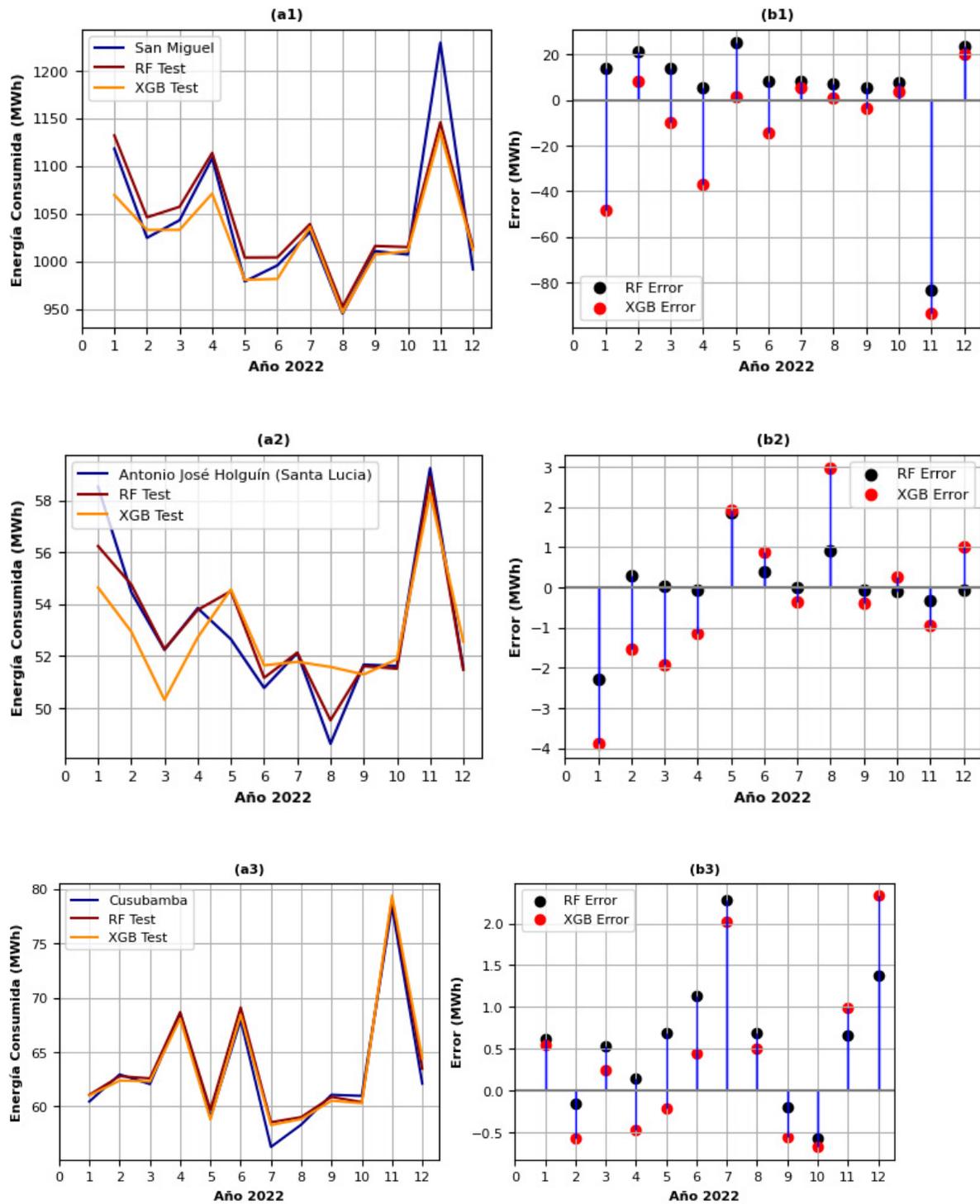
En esta sección, se presentan los resultados detallados sobre el rendimiento de las técnicas Random Forest y XGBoost en la tarea de prever el consumo de energía eléctrica en el cantón Salcedo.

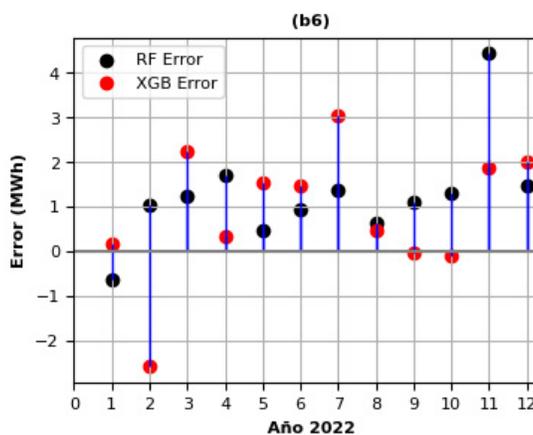
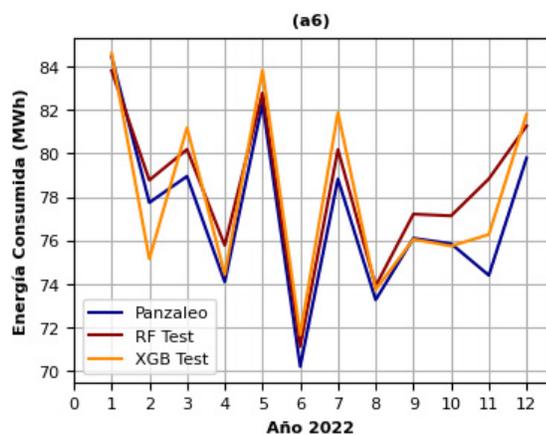
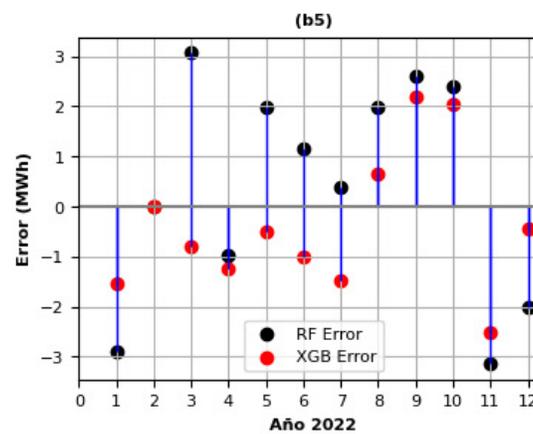
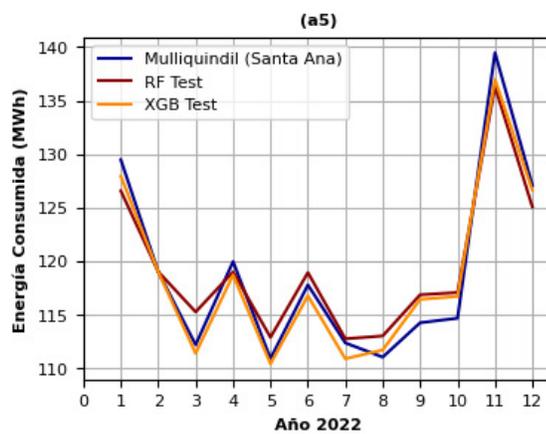
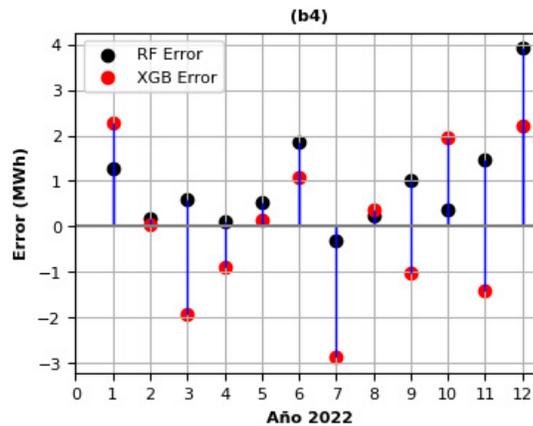
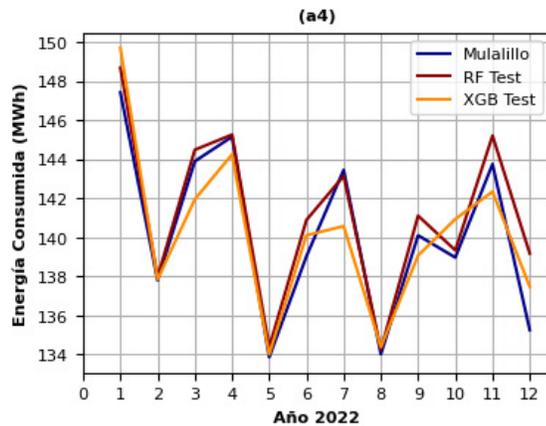
3.1 Resultados de pronóstico

Las Figuras 7(a) presentan el rendimiento de los modelos Random Forest y XGBoost mediante la comparación de las convergencias entre el consumo real y las predicciones generadas para cada una de las parroquias del cantón Salcedo. Por otro lado, las Figuras 7(b) muestran los errores de predicción por mes, proporcionando así una evaluación detallada que facilita la identificación del modelo óptimo para pronosticar el consumo eléctrico.

Figura 7

Rendimiento de Aprendizaje Automático por Parroquia en Salcedo





Una vez completada la metodología de entrenamiento y evaluación, utilizando ambas técnicas de aprendizaje automático, se llevó a cabo una validación conjunta mediante métricas como el error de la raíz cuadrática de la media (RMSE) y el error porcentual absoluto medio (MAPE). Esta validación permitió determinar el modelo más adecuado para predecir con precisión el consumo eléctrico.

La *Tabla 1* presenta una comparación detallada de los mejores resultados de RMSE y MAPE entre los modelos Random Forest y XGBoost para las diferentes parroquias del cantón Salcedo, utilizando los datos de prueba correspondientes al horizonte de predicción al año 2022. Se observa que al evaluar el RMSE y MAPE, se encontró que Random Forest mostró un ajuste más adecuado en la previsión del consumo eléctrico en general.

Tabla 1

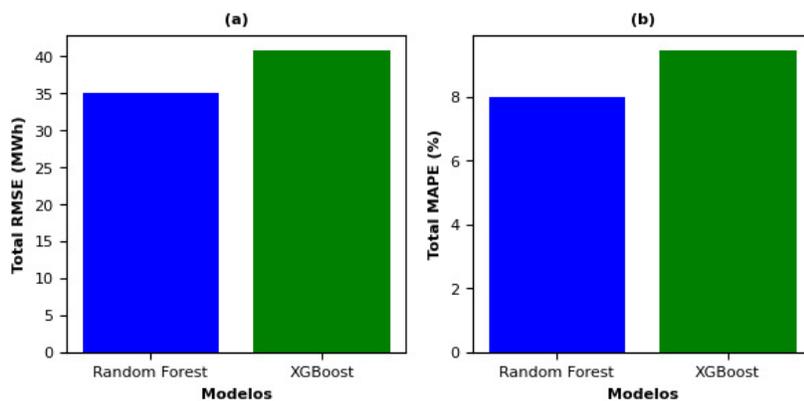
Rendimiento de los Modelos por Grupos de Consumo

Técnicas de aprendizaje automático	Random Forest		XGBoost	
Parroquias	RMSE	MAPE	RMSE	MAPE
San Miguel	27857.75	1.715	33328.98	1.836
Antonio José Holguín (Santa Lucía)	905.21	0.980	1775.84	2.678
Cusubamba	950.22	1.228	1027.47	1.283
Mulalillo	1429.02	0.706	1612.59	0.957
Mulliquindil (Santa Ana)	2135.16	1.576	1411.62	1.001
Panzaleo	1680.45	1.78	1667.40	1.706

En la *Figura 8* se presenta el error absoluto total (RMSE) y el error porcentual total (MAPE), destacando que el Modelo Random Forest mostró un rendimiento superior, mientras que el desempeño del Modelo XGBoost fue inferior en comparación. Asimismo, al analizar los errores porcentuales se observa que Random Forest exhibe un rendimiento superior a XGBoost.

Figura 8

Barras de Errores Absolutos y Porcentuales de los Modelos



Conclusiones

En este estudio se abordó la problemática de la previsión del consumo de energía eléctrica en el cantón Salcedo, Ecuador, dada la creciente demanda generada por el desarrollo global y la expansión de tecnologías. Se reconoció la importancia de contar con proyecciones precisas para la gestión eficiente de sistemas eléctricos, permitiendo a las compañías de servicios públicos planificar de manera más efectiva la generación y distribución de energía.

En respuesta a esta necesidad se exploraron y compararon dos técnicas de aprendizaje automático, Random Forest (RF) y Extreme Gradient Boosting (XGBoost) para la previsión del consumo energético. Estas técnicas demostraron ser valiosas al superar a enfoques tradicionales en escenarios de series temporales impredecibles o inusuales.

La metodología adoptada se basó en la Estructura de Descomposición del Trabajo (WBS), proporcionando una guía clara para la recopilación de datos, el preprocesamiento, el entrenamiento de modelos y la evaluación. Los datos utilizados abarcaron desde enero de 2017 hasta diciembre de 2022, centrándose en clientes residenciales de seis parroquias del cantón Salcedo.

Durante la fase de entrenamiento se seleccionaron cinco predictores claves: año, mes, parroquia, número de clientes y facturación. Ambos modelos, Random Forest y XGBoost, fueron entrenados y evaluados con el objetivo de identificar el mejor rendimiento en la predicción del consumo eléctrico.

Los resultados revelaron que Random Forest exhibió un rendimiento superior en términos de RMSE y MAPE en comparación con XGBoost para todas las parroquias evaluadas en el año 2022. Este hallazgo sugiere que, en el contexto de este estudio, Random Forest podría ser la elección preferida para la previsión del consumo eléctrico en el cantón Salcedo.

Estos resultados tienen implicaciones significativas para la planificación y gestión eficiente de la generación y transmisión de energía eléctrica en la región. Al adoptar técnicas avanzadas de aprendizaje automático, las empresas y autoridades locales pueden mejorar la precisión de las predicciones, contribuyendo así a evitar situaciones críticas como apagones o racionamientos.

Por último, resulta relevante subrayar que tanto Random Forest, como XGBoost, exhibieron una destacada habilidad para detectar patrones de comportamiento atípicos en los conjuntos de datos de entrenamiento en los diversos grupos de consumo. No obstante, se enfatiza la importancia de considerar para futuras investigaciones la incorporación de variables climáticas, factores estacionales, días festivos y tendencias de consumo a lo largo del tiempo. Esta adición potencial de variables podría resultar en mejoras significativas en la precisión de los modelos predictivos.

Referencias

- Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, 203, 810–821. <https://doi.org/10.1016/J.JCLEPRO.2018.08.207>
- ARCERNNR. (2023). Reportes de Información Estadística del Sector Eléctrico. *SISDAT*. <http://reportes.control-recursosyenergia.gob.ec/>
- Banik, R., Das, P., Ray, S., & Biswas, A. (2021). Prediction of electrical energy consumption based on machine learning technique. *Electrical Engineering*, 103(2), 909–920. <https://doi.org/10.1007/S00202-020-01126-Z/METRICS>
- Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied Energy*, 238, 1312–1326. <https://doi.org/10.1016/J.APENERGY.2019.01.113>
- Carrillo, D. B., Pazuña, W. P., & Quinatoa, C. I. (2023). Forecasting Energy Consumption in the Chimborazo Province, Ecuador, Using Random Forest and XGBoost Algorithms. 2023 1st International Conference on Advanced Engineering and Technologies, ICONNIC 2023 - Proceeding, 66–72. <https://doi.org/10.1109/ICONNIC59854.2023.10467276>
- Dhaliwal, S., Nahid, A.-A., & Abbas, R. (2018). Effective Intrusion Detection System Using XGBoost. *Information*, 9(7). <https://doi.org/10.3390/info9070149>
- Guachimboza-Davalos, J. I., Llanes-Cedeño, E. A., Rubio-Aguilar, R., Peralta-Zurita, D., & Núñez-Barrionuevo, O. (2021). Prediction of Monthly Electricity Consumption by Cantons in Ecuador Through Neural Networks: A Case Study. *Advances in Intelligent Systems and Computing*, 1273 AISC, 28–37. https://doi.org/10.1007/978-3-030-59194-6_3/COVER
- Li, R., Jiang, P., Yang, H., & Li, C. (2020). A novel hybrid forecasting scheme for electricity demand time series. *Sustainable Cities and Society*, 55, 102036. <https://doi.org/10.1016/J.SCS.2020.102036>
- Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., Lu, S., Fan, H., Yang, L., & Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, 8, 1087–1095. <https://doi.org/10.1016/J.EGYR.2022.02.251>
- Mañay, E., Chiliquinga, M., Taco, H., y Moreno, M. (2022). Sistema de Internet de las cosas para el monitoreo del índice ultravioleta en la comunidad de Chirinche Bajo. *REVISTA ODIGOS*, 3(2), 9–25. <https://doi.org/10.35290/RO.V3N2.2022.595>

- Mañay, E., Martínez, D., Chiliquinga, M., Criollo, V., Rivera, E., & Toasa, R. (2022). Low-cost deep learning-based architecture for detecting cardiac arrhythmias in ECG signals. *Iberian Conference on Information Systems and Technologies, CISTI, 2022-June*. <https://doi.org/10.23919/CISTI54924.2022.9820243>
- Mir, A., Alghassab, M., Ullah, K., Khan, Z., Lu, Y., & Imran, M. (2020). A Review of Electricity Demand Forecasting in Low and Middle Income Countries: The Demand Determinants and Horizons. *Sustainability, 12*(15), 5931. <https://doi.org/10.3390/SU12155931>
- Rojas, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Risti, E28*, 586–599.
- Rosero-Garcia, J., Llanes-Cedeño, E., Arciniega-Rocha, R., & López-Villada, J. (2021). Analysis of Prediction and Clustering of Electricity Consumption in the Province of Imbabura-Ecuador for the Planning of Energy Resources. *Intelligent Computing - Proceedings of the 2021 Computing Conference*, 1073–1084. https://doi.org/10.1007/978-3-030-80126-7_75/COVER
- Ruiz-Abellón, M. D., Gabaldón, A., & Guillamón, A. (2018). Load Forecasting for a Campus University Using Ensemble Methods Based on Regression Trees. *Energies, 11*(8). <https://doi.org/10.3390/en11082038>
- Shin, S.-Y., & Woo, H.-G. (2022). Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies, 15*(13). <https://doi.org/10.3390/en15134880>



Copyright (2024) © Oscar Fabricio Chicaiza Yugcha, Carlos Julio Martínez Guaman, Israel Antonio Orozco Manobanda, Ángel Danilo Arellano Castro



Este texto está protegido bajo una licencia internacional [Creative Commons](#) 4.0.

Usted es libre para Compartir—copiar y redistribuir el material en cualquier medio o formato — y Adaptar el documento — remezclar, transformar y crear a partir del material—para cualquier propósito, incluso para fines comerciales, siempre que cumpla las condiciones de Atribución. Usted debe dar crédito a la obra original de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios.

Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace de la obra.

[Resumen de licencia](#) – [Texto completo de la licencia](#)