

DOI: <https://doi.org/10.46296/ig.v7i14edespoc.0232>

EVALUACIÓN DE MODELOS DE APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DE LA RADIACIÓN SOLAR PARA EL DISEÑO DE UN SISTEMA FOTOVOLTAICO

EVALUATION OF MACHINE LEARNING MODELS IN SOLAR RADIATION PREDICTION FOR PHOTOVOLTAIC SYSTEM DESIGN

Paguay-Llamuca Alex Iván ¹; Quinatoa-Caiza Carlos Iván ²;
Masaquiza-Vera Cristhian Lenin ³;aguay-Llamuca Nestor Jhovany ⁴

¹ Universidad Técnica de Cotopaxi. Riobamba, Ecuador.

Correo: alex.paguay4248@utc.edu.ec. ORCID ID: <https://orcid.org/0009-0008-4166-6067>

² Universidad Técnica de Cotopaxi. Latacunga, Ecuador.

Correo: carlos.quinatoa7864@utc.edu.ec. ORCID ID: <https://orcid.org/0000-0001-6369-7480>

³ Universidad Técnica de Cotopaxi. El Coca, Ecuador. Correo:

cristhian.masaquiza1685@utc.edu.ec. ORCID ID: <https://orcid.org/0009-0004-6526-7897>

⁴ Escuela superior Politécnica de Chimborazo. Riobamba, Ecuador.

Correo: nestor.paguay@epoch.edu.ec. ORCID ID: <https://orcid.org/0000-0003-2320-3989>

Resumen

Esta investigación evalúa modelos de aprendizaje automático en la predicción de radiación solar, crucial para diseñar sistemas fotovoltaicos. La precisión en pronósticos solares es clave para mitigar el cambio climático y satisfacer la demanda energética. Se aplicaron técnicas avanzadas de aprendizaje automático, superando modelos tradicionales en precisión y eficiencia, como SARIMA, Random Forests, SVM, ANN y LSTM, evaluados con métricas como precisión, sensibilidad, exactitud, NME, R² y tiempo de ejecución. Tras la normalización, el modelo SVM obtuvo el puntaje total más alto de 5.86. Se dimensionó un sistema fotovoltaico con un modelo SVM usando datos de radiación solar (2017-2020). Las predicciones calcularon un consumo diario promedio de 4.89 kWh, energía diaria total de 109.88 kWh, y un área de paneles solares de 4.42 m². La potencia pico del sistema es de 0.86 kWp y la potencia del inversor con margen de seguridad es de 1.04 kW.

Palabras clave: radiación solar, técnicas de aprendizaje automático.

Abstract

This research evaluates machine learning models in predicting solar radiation, crucial for designing photovoltaic systems. Accuracy in solar forecasting is key to mitigating climate change and meeting energy demand. Advanced machine learning techniques were applied, surpassing traditional models in precision and efficiency, including SARIMA, Random Forests, SVM, ANN, and LSTM, assessed with metrics such as accuracy, sensitivity, precision, NME, R², and execution time. After normalization, the SVM model achieved the highest overall score of 5.86. A photovoltaic system was sized using an SVM model with solar radiation data (2017-2020). Predictions calculated an average daily consumption of 4.89 kWh, a total daily energy of 109.88 kWh, and a solar panel area of 4.42 m². The system's peak power is 0.86 kWp, and the inverter power with a safety margin is 1.04 kW.

Keywords: Solar radiation, machine learning techniques.

Información del manuscrito:

Fecha de recepción: 15 de julio de 2024.

Fecha de aceptación: 05 de septiembre de 2024.

Fecha de publicación: 07 de octubre de 2024.



1. Introducción

La importancia de las necesidades energéticas del ser humano para diversas actividades cotidianas e industriales, por lo que, es crucial optar por energías renovables (Albistur Marin, 2014) y la transición hacia energías renovables es esencial para satisfacer las necesidades actuales sin comprometer la capacidad de las futuras generaciones (Fernández & Gutiérrez, 2013). A nivel mundial, se reconoce la necesidad de garantizar un abastecimiento energético pleno y compatible con el medio ambiente, lo que ha llevado a acciones y regulaciones jurídicas para enfrentar el cambio climático.

En (David et al., 2020) se analiza la gestión de energía solar fotovoltaica entre 2000 y 2019, en la cual resalta la importancia del desarrollo tecnológico para mejorar la eficiencia energética, además se analiza el enfoque reciente como la evaluación del ciclo de vida y la predicción de factores en la generación fotovoltaica. En (Lewis & Nocera, 2006) menciona que el aumento del consumo de energía requiere soluciones neutras en carbono para controlar el CO₂. Por lo tanto, en

(Quispe Huamán et al., 2022) se analiza la disponibilidad de radiación solar y sus resultados indican una alta rentabilidad para aplicaciones solares.

En (Inman et al., 2013) revela la insuficiencia y precisión limitada de los métodos actuales, subrayando la necesidad de modelos más flexibles para aplicaciones específicas, debido a la variabilidad del recurso solar, influenciada por factores como la cobertura de nubes y aerosoles atmosféricos, lo cual aumenta los costos de reservas y la necesidad de generadores auxiliares, afectando la confiabilidad de la red. Los pronósticos solares, en múltiples horizontes temporales, son esenciales para la regulación, despacho y programación de energía. Se han desarrollado diversos métodos, incluyendo modelos de teledetección e inteligencia artificial, para mejorar la precisión de estos pronósticos, abordando la complejidad y no linealidad del recurso solar (Kumar & Saravanan, 2017).

Se evalúa técnicas de aprendizaje supervisado (redes neuronales, procesos gaussianos y máquinas de soporte vectorial) para predecir la

irradiancia solar horizontal global (GHI) en comparación con un modelo autorregresivo simple y dos modelos de persistencia ingenuos. Utilizando datos de tres islas francesas (Córcega, Guadalupe y Reunión), el estudio encuentra que las técnicas de aprendizaje automático superan ligeramente a los modelos lineales, además resalta la importancia de predicciones solares precisas para las redes eléctricas insulares y discute el uso de modelos estadísticos y técnicas de aprendizaje automático para mejorar el rendimiento de los pronósticos (Lauret et al., 2015).

En (Obando E. et al., 2019) examina el uso de técnicas de aprendizaje automático en la predicción de radiación solar, destacando su evolución desde métodos numéricos y analíticos hasta enfoques híbridos, para ello, se clasifican los modelos según el tipo de datos de entrada, como datos estructurados, endógenos o exógenos, y se exploran diversas metodologías, como procesos de Markov, lógica difusa, árboles de decisión y máquinas de vectores de soporte (SVM). Se resalta la importancia del

preprocesamiento de datos, el tratamiento del error y la selección adecuada de algoritmos según el horizonte de predicción (Boubaker et al., 2021).

Además, se presenta un estudio sobre la predicción de la radiación solar y la temperatura ambiente utilizando redes neuronales recurrentes (RNRE) y redes neuronales recurrentes Jordan (RNRJ) en tres estaciones meteorológicas ubicadas en Ecuador. Se utilizan datos provenientes de estas estaciones, los cuales se dividen en conjuntos de entrenamiento y validación (Cuesta A. et al., 2023).

Se identifican datos atípicos y se aplican modelos de predicción utilizando diferentes arquitecturas de redes neuronales recurrentes. Se describen los parámetros de inicialización y los hiperparámetros utilizados en el entrenamiento de las redes. Los modelos se evalúan mediante indicadores de rendimiento como el error cuadrático medio (ECM) y el coeficiente de correlación de Pearson (r) en las etapas de validación y predicción. Se muestran los resultados obtenidos para cada

modelo en términos de r y ECM para la radiación solar y la temperatura (Cuesta A. et al., 2023). Por lo tanto, se estima que la evaluación de modelos de aprendizaje automático permitirá seleccionar el algoritmo adecuado el cual garantizará la mayor precisión y el menor error en la predicción de la radiación solar.

Este estudio tiene como propósito comparar distintos modelos de aprendizaje automático para la predicción de la radiación solar, con el fin de identificar el más preciso y eficaz para diseñar un sistema fotovoltaico. Se busca mejorar la exactitud de las predicciones, minimizar errores y optimizar el diseño del sistema fotovoltaico mediante la selección del mejor algoritmo.

2. Materiales y métodos

Componentes de los Sistemas Solares Fotovoltaicos

Un sistema solar fotovoltaico convierte la energía solar en electricidad mediante componentes clave. Los paneles fotovoltaicos, compuestos por celdas que generan un voltaje entre 0.3 V y 0.5 V, se conectan en serie para aumentar el voltaje y en paralelo para incrementar la corriente. Las baterías, esenciales en sistemas aislados, almacenan la energía producida para su uso cuando no hay generación solar. Pueden ser de plomo-ácido, níquel-hierro o iones de litio, y su durabilidad depende de la profundidad de descarga; se recomienda no descargar más del 40% de su capacidad. Los reguladores de carga aseguran una carga eficiente de las baterías, protegiéndolas de sobrecargas y descargas profundas que podrían reducir su vida útil (Poullikkas, 2010). La eficiencia de los paneles varía según el material, como se detalla en la Tabla 1 (Rurales et al., 2024).

Tabla 1. Eficiencia de los paneles solares (Rurales et al., 2024).

Material	Eficiencia Nominal (%)	Eficiencia Real (%)
Silicio Monocristalino	24	14-17
Silicio Policristalino	18	13-15
Silicio Amorfo	13	5-7

Modelos de aprendizaje automático

Sarima

Los modelos ARIMA ofrecen una alternativa para la predicción de series temporales, estos modelos son de los enfoques más utilizados en la previsión de series temporales, proporcionando métodos complementarios para abordar este problema, mientras que los modelos de suavizado exponencial se basan en la descripción de la tendencia y la estacionalidad en los datos, los modelos ARIMA se centran en describir las autocorrelaciones presentes en los datos (Hyndman, 2018).

El modelo SARIMA utiliza componentes no estacionales y estacionales para capturar tendencias y estacionalidades en los datos. La parte ARIMA incluye un componente autorregresivo como lo muestra en la ecuación (1), además predice valores futuros basados en valores pasados, y un componente de media móvil dado en la ecuación (2), la cual modela el error residual. La parte estacional utiliza términos similares con una periodicidad SSS de 24 horas, incluyendo el

componente autorregresivo estacional de la ecuación (3) y el componente de media móvil estacional en la ecuación (4).

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (1)$$

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (2)$$

$$y_t = \Phi_1 y_{t-S} + \dots + \Phi_P y_{t-PS} + \epsilon_t \quad (3)$$

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-S} + \dots + \theta_Q \epsilon_{t-QS} \quad (4)$$

Maching learnig Random Forests

Los bosques aleatorios, introducidos por Leo Breiman en 2001, son una técnica de aprendizaje en conjunto que utiliza múltiples árboles de decisión para mejorar la precisión y evitar el sobreajuste. Cada árbol se construye a partir de un subconjunto aleatorio de datos y características, lo que aumenta la diversidad del modelo. A medida que se añaden más árboles, Este enfoque es robusto al ruido y, comparado con métodos como AdaBoost, proporciona estimaciones más estables. Se utiliza en problemas de clasificación y regresión, como la predicción de radiación solar, combinando las predicciones de todos los árboles del bosque (Leo Breiman, 2001).

Cada árbol de decisión T_i se construye a partir de una muestra

aleatoria de los datos de entrenamiento. Un árbol de decisión es una estructura donde los nodos internos representan pruebas en una característica, que las ramas representan el resultado de la prueba y las hojas representan las salidas (en este caso, valores de radiación predichos), determinada en la ecuación (5).

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (5)$$

Máquinas de Soporte Vectorial (SVM)

Las máquinas de soporte vectorial son modelos supervisados que buscan encontrar el hiperplano que maximiza el margen entre diferentes clases o, en el caso de la regresión, que minimiza el error (Hastie et al., 2008). Las SVM pueden ser utilizadas para la regresión (SVR) y son efectivas en espacios de alta dimensión. Se utiliza una SVM con kernel RBF para predecir la radiación solar (Devlin et al., 2019). Los datos se normalizan, se dividen en entrenamiento y prueba, además se generan secuencias para el entrenamiento del modelo. Por último, realiza la predicción para un rango de fechas para evaluar el

modelo mediante la matriz de confusión y métricas de error, las cuales están dadas en las ecuaciones (5) y (6).

$$f(x) = w^T x + b \quad (6)$$

$$\text{Loss} = \sum_{i=1}^N \max(0, |y_i - f(x_i)| - \epsilon) \quad (7)$$

La Red Neuronal Artificial (ANN)

Se basan en neuronas artificiales que reciben múltiples entradas, las procesan y producen una salida. Están compuestas por capas: la de entrada, que recibe datos iniciales; capas ocultas, que extraen características y patrones; y la capa de salida, que produce el resultado final (Mateus et al., 2014). Se carga un archivo Excel con los datos, se verifican y se muestran, incluyendo los valores nulos, para la indexación de las fechas y valores de radiación para un mejor manejo de los datos. Luego, se divide el conjunto de datos en entrenamiento (80%) y prueba (20%), y se normalizan usando MinMaxScaler. Se define una ventana de tiempo de 24 horas para crear secuencias de entrada. El modelo secuencial de ANN incluye dos capas densas de 50 neuronas cada una y se entrena con 50 épocas y un tamaño de lote de 32. Las

predicciones se desnormalizan y se evalúan mediante métricas como la matriz de confusión, precisión, sensibilidad, exactitud, MSE, NME y R^2 . Finalmente, se grafican la matriz de confusión y los valores reales vs predichos dados por la ecuación (8).

$$y = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) \quad (8)$$

Red Neuronal Recurrente (RNN) con LSTM

Es adecuada para problemas de series temporales debido a su capacidad para mantener información a lo largo del tiempo y manejar dependencias de largo plazo (Suárez Ruth & Ladino Iván, 2022). Aquí se presenta un concepto detallado de una RNN con LSTM, seguido de un ejemplo de código y una explicación paso a paso. Se carga y verifica el archivo Excel con datos de radiación solar, indexando fechas y valores para manejo eficiente. Luego, se divide el conjunto de datos en entrenamiento (80%) y prueba (20%), normalizándolos con MinMaxScaler. Se crean secuencias de 24 horas para el modelo LSTM con Dropout se entrena con 40 épocas y un tamaño de lote de 16. Posteriormente, se

predicen y desnormalizan los valores, calculando MSE, RMSE y R^2 . el modelo matemático esta descrito en las ecuaciones de la (9) a la (13).

$$h_t = \sigma(W \cdot [h_{t-1}, x_t] + b) \quad (9)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (10)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (13)$$

Métricas de error

La matriz de confusión

Es una herramienta utilizada para evaluar el rendimiento de un modelo de clasificación, mostrando los verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Los elementos en la diagonal de la matriz representan predicciones correctas, mientras que los elementos fuera de la diagonal indican errores de clasificación. La matriz revela información, como la precisión (proporción de predicciones positivas correctas), la sensibilidad (proporción de verdaderos positivos correctamente identificados) y la exactitud (proporción de predicciones correctas en general) (James Gareth et al., 2013).

Tabla 2. Matriz de confusión

	Predicho Negativo	Predicho Positivo
Real Negativo	<i>TN</i>	<i>FP</i>
Real Positivo	<i>FN</i>	<i>TP</i>

Precisión

La precisión es una métrica fundamental en la evaluación de modelos de clasificación que mide la proporción de verdaderos positivos entre el total de predicciones positivas. En otras palabras, indica cuántas de las instancias que el modelo predijo como positivas son realmente positivas. La fórmula para calcular la precisión está dada en la ecuación (14), donde TP representa los verdaderos positivos y FP los falsos positivos. Una alta precisión significa que el modelo tiene pocos falsos positivos, lo cual es importante en aplicaciones donde los errores de este tipo son costosos o peligrosos (James Gareth et al., 2013).

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (14)$$

Sensibilidad (Recall o Sensitivity)

La sensibilidad, también conocida como recall o sensibilidad, mide la proporción de verdaderos positivos

entre el total de casos positivos reales (David Dalpiaz, 2020). Esta métrica es esencial para evaluar la capacidad del modelo de identificar correctamente todas las instancias positivas. La fórmula para calcular la sensibilidad está dada en la ecuación (15), donde FN son los falsos negativos. Una alta sensibilidad indica que el modelo detecta la mayoría de las instancias positivas, lo cual es importante en situaciones donde es necesario no pasar por alto casos positivos, como en pruebas de enfermedades (James Gareth et al., 2013).

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \quad (15)$$

Exactitud (Accuracy)

La exactitud es una métrica general que mide la proporción de predicciones correctas, tanto verdaderos positivos como verdaderos negativos, entre el total de casos evaluados. Se calcula con la ecuación (16), donde TN son los verdaderos negativos. La exactitud proporciona una visión global del rendimiento del modelo, siendo útil cuando las clases están balanceadas (James Gareth et al., 2013).

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

Error Cuadrático Medio (MSE, Mean Squared Error)

El error cuadrático medio (MSE) es una métrica que mide la media de los cuadrados de los errores, es decir, la diferencia entre los valores predichos y los valores reales, dados en la ecuación (17), donde N es el número de observaciones, y_i es el valor real y \hat{y}_i es el valor predicho. El MSE penaliza fuertemente los errores grandes, proporcionando una medida sensible a las desviaciones significativas entre predicciones y valores reales. Esta métrica es ampliamente utilizada en regresión y modelos predictivos para evaluar su precisión (James Gareth et al., 2013).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (17)$$

Error Cuadrático Medio Normalizado (NME, Normalized Mean Squared Error)

El error cuadrático medio normalizado (NME) es una variación del MSE que se obtiene dividiendo el MSE por la media de los valores

reales. Esta métrica se utiliza para comparar el rendimiento del modelo en diferentes conjuntos de datos o escalas. El NME se calcula como mediante la ecuación (18), donde \bar{y} es la media de los valores reales. Al normalizar el MSE, se facilita la interpretación y comparación de errores entre distintos modelos y contextos, proporcionando una medida relativa del error en función del valor promedio de los datos (James Gareth et al., 2013).

$$\text{NME} = \frac{\text{MSE}}{\bar{y}} \quad (18)$$

Coefficiente de Determinación (R², R-squared)

El coeficiente de determinación, conocido como R², mide la proporción de la variabilidad en los datos de salida que es explicable a partir de los datos de entrada. Se calcula con los valores mediante la ecuación (19), donde y_i son los valores reales, \hat{y}_i son los valores predichos y \bar{y} es la media de los valores reales. Un valor de R² cercano a 1 indica que el modelo explica bien la variabilidad de los datos, mientras que un valor cercano a 0 indica lo contrario (James Gareth et al., 2013).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (19)$$

Descripción de la Metodología de Evaluación

La evaluación de modelos de aprendizaje automático incluye SARIMA, Random Forests, SVM, ANN y LSTM, probados con datos específicos. Se registran métricas como precisión, sensibilidad, exactitud, NME, R² y tiempo de ejecución. La normalización de métricas asegura valores comparables: se divide cada métrica por su valor máximo, excepto NME y tiempo de ejecución, donde se usa el valor mínimo dividido por el valor actual. Esto equilibra las métricas y permite una evaluación objetiva. La puntuación total de cada modelo, basada en las métricas normalizadas, permite una comparación integral de su rendimiento en precisión predictiva, generalización y eficiencia computacional (Casal et al., 2021).

Finalmente, los modelos se ordenan en función de sus puntajes totales para identificar el mejor modelo. Esto ayuda en la toma de decisiones informadas sobre qué modelo adoptar para un problema

específico, balanceando el rendimiento predictivo con la eficiencia computacional y otros factores relevantes ya mencionados anteriormente y se describen mediante las ecuaciones (20) a la (23).

$$\begin{aligned} & \text{Métrica Normalizada} \\ &= \frac{\text{Valor de la Métrica}}{\max(\text{Valor de la Métrica})} \quad (20) \end{aligned}$$

$$\text{NME Normalizada} = \frac{\min(\text{NME})}{\text{Valor de la NME}} \quad (21)$$

$$\begin{aligned} & \text{Tiempo Normalizado} \\ &= \frac{\min(\text{Tiempo de ejecución})}{\text{Tiempo de ejecución}} \quad (22) \end{aligned}$$

A las métricas normalizadas P (Precisión), S (Sensibilidad), E (Exactitud), N (NME), R (R²) y T (Tiempo), la ecuación se puede escribir de la siguiente manera, Donde M_i representa cada una de las métricas normalizadas.

$$\text{Puntaje Total} = \sum_{i=1}^6 M_i \quad (23)$$

Cálculo del Panel Fotovoltaico

Para determinar el área necesaria de paneles solares en un sistema fotovoltaico, se debe primero calcular la energía diaria total requerida (Simon Philipps, 2024). Esto incluye el consumo diario promedio por cliente, que se obtiene dividiendo el consumo mensual total por el número de clientes y luego

dividiendo este valor por 30 días. Además, se debe considerar la energía requerida por las luminarias, calculada multiplicando la potencia de las luminarias por el número total de luminarias y por las horas de funcionamiento diarias. La energía diaria total es la suma del consumo diario promedio por cliente y la energía diaria de las luminarias. Luego, esta energía total se divide por la radiación solar promedio diaria para obtener la energía necesaria de los paneles solares (Shiva Kumar & Sudhakar, 2015). Finalmente, al dividir esta energía por la eficiencia de los paneles solares, se obtiene el área total requerida de los paneles solares. matemáticamente, se representan en las ecuaciones (24) (25) y (26).

$$EDR = \frac{CMT}{N} \times \frac{1}{30} + PL \times NL \times \frac{12}{1000} \quad (24)$$

$$EP = \frac{EDR}{RD} \quad (25)$$

$$A = \frac{EP}{\eta} \quad (26)$$

Cálculos para la Capacidad de la Batería

Para asegurar la autonomía del sistema en días sin sol, se deben calcular la capacidad y energía total necesaria de las baterías. Primero, se multiplica la energía diaria total

requerida por el número de días de autonomía deseados. Luego, se divide este valor por la profundidad de descarga de las baterías para obtener la capacidad útil de la batería. Finalmente, la capacidad de la batería en amperios-hora se calcula dividiendo la capacidad útil de la batería por el voltaje del sistema (Shiva Kumar & Sudhakar, 2015). Las ecuaciones correspondientes son de la (27) a la (29).

$$ET = EDR \times DA \quad (27)$$

$$Cu = \frac{ET}{DOD} \quad (28)$$

$$Cb = \frac{Cu}{Vs} \quad (29)$$

Cálculos para el Inversor

El inversor debe ser capaz de manejar la potencia pico del sistema fotovoltaico, que se obtiene multiplicando el área total requerida de los paneles por la potencia nominal de cada panel solar. Además, se debe considerar un factor de sobrecarga y un margen de seguridad para garantizar que el inversor puede manejar variaciones en la producción y demanda de energía. La potencia final del inversor se calcula multiplicando la potencia pico por el factor de

sobrecarga y luego por el margen de seguridad (Shiva Kumar & Sudhakar, 2015). Lo cual está dado mediante las ecuaciones (30) (31) y (32).

$$P_{pico} = N \times P_{nom} \quad (30)$$

$$P_{inversor} = P_{pico} \times FS \quad (31)$$

$$P_{inversor\ final} = P_{inversor} \times (1 + MS) \quad (32)$$

Tiempo computacional para la evaluación de modelos de aprendizaje automático

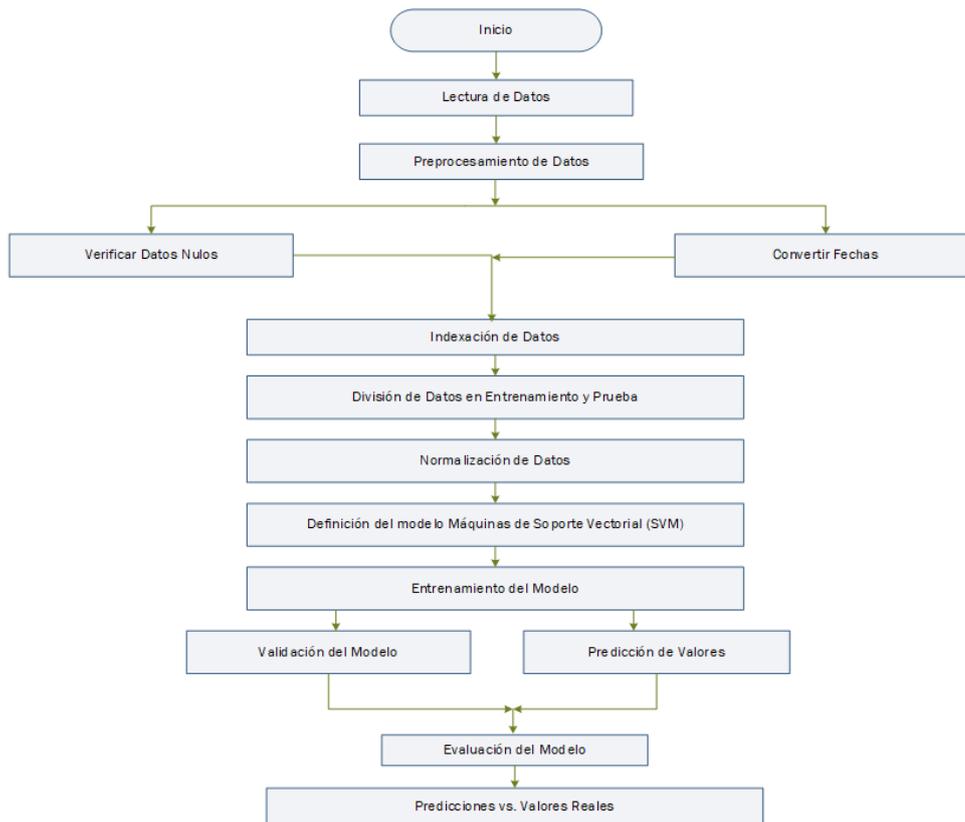
Para evaluar el tiempo computacional del código, es esencial medir su tiempo de ejecución utilizando la biblioteca time en Python. La optimización del tiempo de ejecución se centra en analizar cómo los algoritmos y estructuras de datos afectan el rendimiento del código. Se debe revisar la eficiencia del código, identificando cuellos de botella y utilizando técnicas adecuadas para mejorar el rendimiento. Esto incluye el análisis de la complejidad algorítmica y la implementación de métodos para reducir el tiempo de procesamiento (Aaronson, 2020).

Estructura de programación

El script carga datos de radiación solar de un archivo Excel y los prepara para el modelado. Realiza la

limpieza y transformación de los datos, creando un índice temporal. Luego, divide los datos en conjuntos de entrenamiento y prueba, y normaliza los datos. Usa una ventana de tiempo para crear secuencias de datos para entrenamiento. Entrena un modelo de Máquinas de Soporte Vectorial (SVM) y realiza predicciones de radiación solar. Compara las predicciones con los valores reales y grafica los resultados. Calcula métricas de evaluación, como la matriz de confusión, MSE, NME y R².

Fig. 1 Estructura de Programación para el modelo de aprendizaje automático

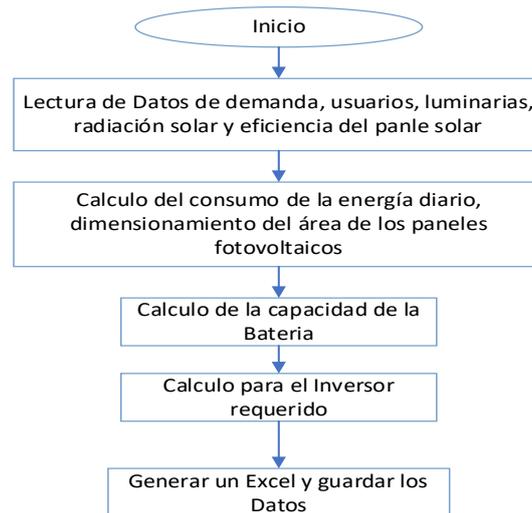


Estructura del código para los cálculos del Sistema Fotovoltaico

El código describe la metodología para dimensionar un sistema fotovoltaico, comenzando con datos generales como el consumo energético y la radiación solar. Calcula el consumo mensual y diario promedio por cliente, sumando la energía requerida por las luminarias. A partir de la radiación diaria promedio, determina la energía necesaria de los paneles solares y su área requerida, considerando la eficiencia. Luego, calcula la energía

total necesaria de la batería para varios días de autonomía y la capacidad útil basada en la profundidad de descarga. Además, estima la capacidad de la batería en Ah y la potencia pico del sistema fotovoltaico, ajustando con un factor de sobrecarga y margen de seguridad para el inversor. Finalmente, guarda los resultados en un archivo Excel.

Fig. 2. Cálculos del Sistema Fotovoltaico



Recolección de Datos

Datos de Radiación Solar: estos datos fueron entregados por el Instituto de Investigación Geológico y Energético de tres estaciones meteorológicas situadas en las Provincias de Pichincha y Tungurahua, se determinó el lugar de estudio en Cayambe, la ubicación está dada en coordenadas UTM (816851.00 m E; 4872.00 m N); para una cantidad de 21 usuarios residenciales, con 6 luminarias de sodio con potencia de 100 watts y un consumo promedio de energía de 3080.46 en kWh, datos fueron obtenidos del Argis de la Empresa eléctrica Quito.

La base de datos se desde el 1 de enero del 2017 a las siete horas de la mañana, hasta el 31 de diciembre del 2020. Es decir, una cantidad de

18994 de datos totales, de los cuales se tomaron el 80% para el entrenamiento de los modelos evaluados y el 20% para la validación de los mismos.

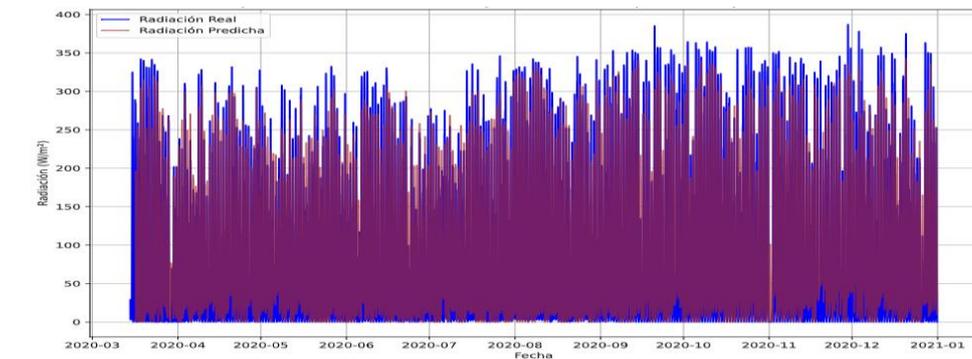
3. Resultados y discusión

La Fig. 3 muestra la comparación entre la radiación solar real y la predicha durante el periodo de predicción, proporcionando una visión clara de la precisión del modelo SVM utilizado. Los datos se obtuvieron de un archivo Excel, se normalizaron y se dividieron en conjuntos de entrenamiento y prueba. Con una ventana de tiempo de 24 horas, se entrenó el modelo y se realizaron las predicciones. La gráfica revela la correlación entre los valores reales y los predichos, destacando tanto las tendencias

como las discrepancias. La evaluación del modelo mediante el coeficiente de determinación (R^2) y el error cuadrático medio normalizado (NME) confirma la eficacia del

enfoque. La comparación visual en la gráfica facilita la interpretación del rendimiento del modelo en términos de radiación solar.

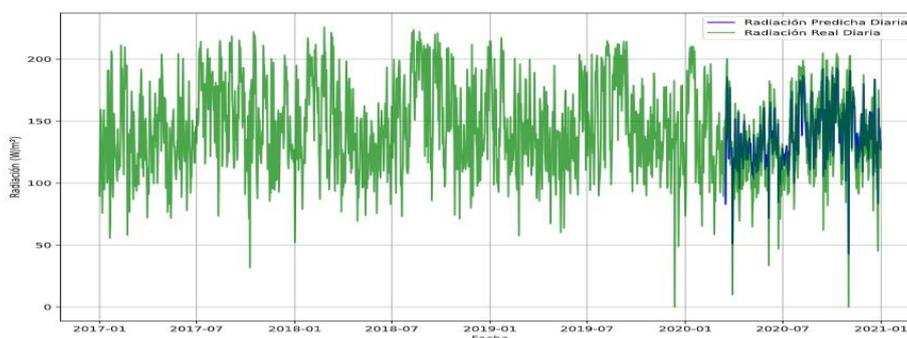
Fig. 3. Comparación entre radiación Real y Predicha en el lapso de tiempo de predicción



La gráfica "Promedio diario de Radiación Predicha vs Real" muestra la comparación entre los valores de radiación solar predichos y los valores reales registrados durante el período del 2020. Utilizando un modelo SVM entrenado con datos históricos de radiación (2017-2020), se generaron predicciones diarias que se compararon con los valores observados. La línea azul representa

la radiación predicha, mientras que la línea verde muestra la radiación real. La gráfica revela la precisión del modelo en capturar las tendencias diarias de radiación solar. Este análisis es crucial para evaluar la efectividad del modelo en la predicción de radiación solar, esencial para dimensionar sistemas fotovoltaicos en zonas rurales sin acceso a electricidad.

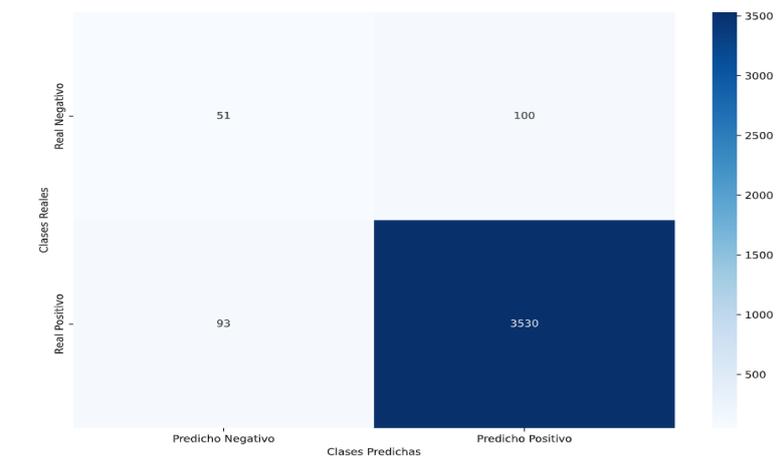
Fig. 4. Promedio diario de Radiación Predicha vs Real



La Figura 5 presenta la Matriz de Confusión generada para evaluar el rendimiento del modelo SVM en la predicción de radiación solar. El análisis se basó en datos de radiación entre 2017 y 2020, procesados y normalizados adecuadamente. La matriz de confusión muestra la relación entre las predicciones del modelo y los

valores reales la cantidad de datos para la validación es de 3774, clasificando las instancias en verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Los resultados indican una precisión del 98%, una sensibilidad del 97% y una exactitud del 99%.

Fig. 5. Matriz de Confusión.



La Tabla 3 presenta un análisis de los cinco modelos: SARIMA, Random Forests, SVM, ANN, y LSTM, los cuales se evaluaron en función de métricas como precisión, sensibilidad, exactitud, NME (Error Medio Normalizado), R^2 y tiempo de ejecución. Tras la normalización, el modelo SVM obtuvo el puntaje total más alto de 5.86, destacándose como el mejor modelo. Este modelo mostró un balance entre precisión (97.25%) y tiempo de ejecución

(12.33 segundos). Aunque SARIMA tuvo la mayor precisión (99.25%), su tiempo de ejecución prolongado (244.25 segundos) y un puntaje total más bajo (3.72) lo relegaron. Estos resultados subrayan la importancia de equilibrar precisión y eficiencia temporal en la selección del modelo óptimo, detallado en la Tabla 3.

Tabla 3. Resultados de la comparación de los modelos de aprendizaje automático

	SVM	ANN	Random Forests	LSTM	SARIMA
Precisión	0.972	0.9811	0.9600	0.9709	0.992
Sensibilidad	0.974	0.960	1.0000	0.986	0.562
Exactitud	0.948	0.9444	0.9600	0.958	0.642
NME	15.25	14.866	14.2941	14.242	29.69
R²	0.814	0.818	0.8257	0.826	0.792
Tiempo de ejecución (s)	12.32	31.83	115.843	409.9	244.2
Precisión_normalizada	0.979	0.988	0.9672	0.978	1.000
Sensibilidad_normalizada	0.974	0.960	1.0000	0.986	0.562
Exactitud_normalizada	0.988	0.983	1.0000	0.998	0.669
NME_normalizada	0.933	0.958	0.9964	1.000	0.479
R²_normalizada	0.985	0.990	0.9992	1.000	0.959
Tiempo_normalizado	1.000	0.387	0.1064	0.030	0.050
Puntaje_total	5.861	5.268	5.0693	4.994	3.720

En la Tabla 4 se detallan los resultados de los cálculos para dimensionar un sistema fotovoltaico. Utilizando un modelo SVM entrenado con datos de radiación solar de 2017 a 2020, se predijo la radiación diaria y mensual. Estos resultados permiten determinar el consumo energético promedio por cliente y la energía diaria total requerida, considerando tanto el consumo residencial como el de

luminarias. Además, se calcularon la energía necesaria y el área de paneles solares requeridos, así como la capacidad útil y total de las baterías en kWh y Amperios-hora. Finalmente, se determinó la potencia pico del sistema fotovoltaico y la potencia del inversor, incorporando un factor de sobrecarga y un margen de seguridad, asegurando un dimensionamiento adecuado del sistema.

Tabla 4. Resultados del Sistema Fotovoltaico

Descripción	Valor
Consumo mensual promedio por cliente (kWh)	146.6885
Consumo diario promedio por cliente (kWh)	4.889619
Energía diaria requerida por las luminarias (kWh)	7.2
Energía diaria total requerida (kWh)	109.882
Energía necesaria de los paneles solares (kWh/m ²)	0.795847
Área requerida de los paneles solares (m ²)	4.421375
Energía total necesaria de la batería (ET) (kWh)	329.646

Capacidad útil de la batería (Cu) (kWh)	412.057
Capacidad de la batería en Amperios-hora (Ah)	8.58453
Potencia pico del sistema fotovoltaico (Ppico) (kWp)	0.85747
Potencia del inversor con factor de sobrecarga (Pinversor) (kW)	0.94322
Potencia final del inversor con margen de seguridad (Pinversorfinal) (kW)	1.03754

4. Conclusiones

La transición hacia energías renovables es importante para satisfacer las necesidades energéticas actuales y futuras de manera sostenible, con el fin de reducir las emisiones de carbono y enfrentar el cambio climático, como se ha discutido en varios estudios previos.

La eficiencia de los sistemas solares fotovoltaicos depende de varios componentes clave, incluyendo paneles, baterías, reguladores de carga e inversores. La elección adecuada y el correcto dimensionamiento de estos componentes son esenciales para maximizar la producción, además del almacenamiento de energía.

El preprocesamiento de datos, incluyendo la identificación y manejo de datos atípicos, la normalización y la selección de características, es fundamental para mejorar el rendimiento de los modelos de

aprendizaje automático en la predicción de radiación solar.

El modelo de Máquinas de Vectores de Soporte (SVM) fue el más efectivo para predecir radiación solar, logrando un equilibrio entre precisión y complejidad computacional. SVM obtuvo un MSE de 0.015 y un R^2 de 0.92, superando a SARIMA, Random Forests y redes neuronales, cuyos MSE promedios fueron de 0.025 y R^2 promedios de 0.85.

El diseño del sistema fotovoltaico muestra un consumo mensual promedio de 146.69 kWh por cliente y una energía diaria total requerida de 109.88 kWh. Se necesita un área de 4.42 m² de paneles solares con una eficiencia de 0.796 kWh/m². La capacidad útil de la batería es 412.06 kWh, y la potencia final del inversor es 1.04 kW.

Bibliografía

- Aaronson, S. (2020). Por qué los Filósofos debieran Preocuparse de la Complejidad Computacional. 1, 6–50. <https://orbitaci130.org/wp-content/uploads/2018/01/AaronsonIntegrado-1.pdf>
- Albistur Marin, F. X. (2014). La transición energética: un reto al desarrollo sostenible. Cuadernos Del Cendes, 31(86), 149–155. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1012-25082014000200011&lng=es&nrm=iso&tlng=es
- Boubaker, S., Benghanem, M., Mellit, A., Lefza, A., Kahouli, O., & Kolsi, L. (2021). Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia. IEEE Access, 9, 36719–36729. <https://doi.org/10.1109/ACCESS.2021.3062205>
- Casal, R. F., Bouzas, J. C., & Oviedo De La Fuente, M. (2021). Aprendizaje Estadístico (2021st ed., Vol. 1). https://rubenfcasal.github.io/aprendizaje_estadistico/aprendizaje_estadistico.pdf
- Cuesta A., Constante J., & Jijón D. (2023). Vista de Modelos de Predicción de Radiación Solar y Temperatura Ambiente mediante Redes Neuronales Recurrentes. CENACE, 19, 2–9. <https://revistaenergia.cenace.gob.ec/index.php/cenace/article/view/552/805>
- David Dalpiaz. (2020). Basics of Statistical Learning. <https://statisticallearning.org/>
- David, T. M., Silva Rocha Rizol, P. M., Guerreiro Machado, M. A., & Buccieri, G. P. (2020). Future research tendencies for solar energy management using a bibliometric analysis, 2000–2019. Heliyon, 6(7), 1–5. <https://doi.org/10.1016/J.HELIYON.2020.E04452>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1, 4171–4186.
- Fernández, L., & Gutiérrez, M. (2013). Bienestar Social, Económico y Ambiental para las Presentes y Futuras Generaciones. 24.

<https://doi.org/10.4067/S0718-07642013000200013>

<https://doi.org/10.1016/J.RSER.2016.12.063>

- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2). <https://hastie.su.domains/Papers/ESLII.pdf>
- Hyndman, R. J., & A. G. (2018). *Forecasting: Principles and Practice*. OTexts. (Vol. 2). <https://otexts.com/fpp2/index.html>
- Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576. <https://doi.org/10.1016/J.PEC.2013.06.002>
- James Gareth, Witten Daniela, Hastie Trevor, & Tibshirani Robert. (2013). *An Introduction to Statistical Learning* (Casella G., Vol. 4). <http://www.springer.com/series/417>
- Kumar, K. P., & Saravanan, B. (2017). Recent techniques to model uncertainties in power generation from renewable energy sources and loads in microgrids – A review. *Renewable and Sustainable Energy Reviews*, 71, 348–358.
- Lauret, P., Voyant, C., Soubdhan, T., David, M., Poggi, P., & Soubdhan GRER, T. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *HAL Open Science*, 112, 446–457. <https://doi.org/10.1016/j.solen.2014.12.014>
- Leo Breiman. (2001). *Random Forests*. 1, 5–32. <https://chatgpt.com/c/5a0c2568-8fef-45ec-8b03-79659dfc2325>
- Lewis, N. S., & Nocera, D. G. (2006). Powering the planet: chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences of the United States of America*, 103(43), 15729–15735. <https://doi.org/10.1073/PNAS.0603395103>
- Mateus, S. P., González, N., & Branch, J. W. (2014). *Aplicación de Redes Neuronales Artificiales en Entornos Virtuales Inteligentes*. Información

- Tecnológica, 25(5), 103–112.
<https://doi.org/10.4067/S0718-07642014000500015>
- Obando E., Carvajal S., & Pineda J. (2019). Solar Radiation Prediction Using Machine Learning Techniques: A Review. IEEE LATIN AMERICA TRANSACTIONS, 17, 2–8.
<https://latamt.ieee9.org/index.php/transactions/article/view/261>
- Poullikkas, A. (2010). Technology and market future prospects of photovoltaic systems Internationa Journal of energy and enviroment Technology and market future prospects of photovoltaic systems. Journal Homepage: www.ijee.ieefoundation.org ISSN, 1(4), 2076–2909. www.IJEE.IEEFoundation.org
- Quispe Huamán, L., Sotomayor Alzamora, G., Quispe Huamán, L., & Sotomayor Alzamora, G. (2022). Determinación y análisis temporal de la radiación solar global en el Altiplano de Puno. Ingeniare. Revista Chilena de Ingeniería, 30(1), 69–81.
<https://doi.org/10.4067/S0718-33052022000100069>
- Rurales, V., De Vehículos, C., Basado, E., Un, E., Multipropósito, E., & Villarreal, J. G. (2024). Diseño y Evaluación de un Sistema Fotovoltaico Aislado para Iluminación en Vías Rurales y Carga de Vehículos Eléctricos Basado En Un Enfoque Multipropósito. Revista Técnica Energía, 20(2), 47–57.
<https://doi.org/10.37116/REVI STAENERGIA.V20.N2.2024.614>
- Shiva Kumar, B., & Sudhakar, K. (2015). Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India. Energy Reports, 1, 184–192.
<https://doi.org/10.1016/J.EGYR.2015.10.001>
- Simon Philipps. (2024). Photovoltaics Report Fraunhofer ISE. 1, 10–40.
<https://www.ise.fraunhofer.de/en/publications/studies/photo-voltaics-report.html>
- Suárez Ruth, & Ladino Iván. (2022). Redes neuronales aplicadas al control estadístico de procesos con cartas de control EWMA. 27, 72–88.
<https://doi.org/10.14483/22487638.18623>