REVISTA CIENTÍFICA CONECTIVIDAD

Conectividad
ISSN: 2806-5875
ISSN-L: 2806-5875
revista@ister.edu.ec
Tecnológico Superior Rumiñahui
Ecuador

alomo, Wilson; Quinatoa, Carlos; Mullo, Mauricio; Castillo, Jessica N
Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático
Conectividad, vol. 6, núm. 1, 2025, Enero-Junio, pp. 338-355
Tecnológico Superior Rumiñahui
Sangolquí, Ecuador

DOI: https://doi.org/10.37431/conectividad.v6i1.196

Disponible en: https://www.redalyc.org/articulo.oa?id=777882650028

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org
Sistema de Información Científica Redalyc
Red de revistas científicas de Acceso Abierto diamante
Infraestructura abierta no comercial propiedad de la academia

REVISTA CIENTÍFICA
CONECTIVIDAD
Instituto Superior Tecnológico Universitario Rumiñahui

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

# *Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático*

## *Solar Radiation Forecasting in Photovoltaic Systems Utilizing Aprendizaje Automático Techniques*

Wilson Palomo[1], Carlos Quinatoa[2], Mauricio Mullo[3], Jessica N. Castillo[4]

[1] Universidad Técnica de Cotopaxi, wilson.palomo1@utc.edu.ec, Latacunga, Ecuador
[2] Universidad Técnica de Cotopaxi, carlos.quinatoa4@utc.edu.ec, Latacunga, Ecuador
[3] Universidad Técnica de Cotopaxi, mauricio.mullo@utc.edu.ec, Latacunga, Ecuador
[4] Universidad Técnica de Cotopaxi, jessica.castillo@utc.edu.ec, Latacunga, Ecuador

Autor para correspondencia: wilson.alomo1@utc.edu.ec

## RESUMEN

Esta investigación responde a la creciente demanda de energías renovables, enfocándose específicamente en los sistemas fotovoltaicos que aprovechan la energía solar como una solución viable y sostenible. La metodología implementada incluyó el análisis y tratamiento de los datos de radiación solar recopilados cada hora durante el periodo 2017-2023. Estos datos fueron fundamentales para realizar las predicciones de enero, febrero y marzo del 2024. El propósito de estas predicciones fue optimizar el dimensionamiento de un sistema fotovoltaico apropiado para un área urbana. Para este propósito, se utilizó un algoritmo de árbol de decisión, una técnica destacada dentro del campo del aprendizaje automático, implementada mediante el software Python por su facilidad de acceso y versatilidad. Los resultados se almacenaron en un archivo .xlsx, lo que simplificó el proceso de dimensionamiento del sistema. Además, se incorporaron cálculos de desviación estándar para estimar la radiación solar en los próximos tres meses, permitiendo así un cálculo preciso y adecuado del sistema fotovoltaico necesario. En conclusión, el sistema fotovoltaico diseñado se dimensiono eficazmente a partir del análisis predictivo proporcionado por el algoritmo. Con una potencia pico de 1,26 kWp y una configuración de almacenamiento bien adaptada, este sistema está equipado para cumplir con las demandas energéticas diarias de 123,5 kWh.

**Palabras clave**: Energía renovable, Sistemas fotovoltaicos, Predicción, Árbol de decisión.

## ABSTRACT

This research responds to the growing demand for renewable energy, focusing specifically on photovoltaic systems that harness solar energy as a viable and sustainable solution. The methodology implemented included the analysis of hourly solar radiation data collected during the period 2017-2023. These data were fundamental to make predictions and validate the algorithm used. The purpose of these predictions was to optimize the sizing of a photovoltaic system appropriate for an urban area. For this purpose, a decision tree algorithm was used, a leading technique in the field of Machine Learning, implemented using Python software for its ease of access and versatility. The results were stored in an .xlsx file, which simplified the system sizing process. In addition, standard deviation calculations were incorporated to estimate the solar radiation over the next three months, thus allowing an accurate and adequate calculation of the required PV system. In conclusion, the designed PV system was efficiently sized based on the predictive analysis provided by the algorithm. With a peak power of 1.26 kWp and a well-adapted storage configuration, this system is equipped to meet the daily energy demands of 123.5 kWh.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Key words**: Renewable energy, Photovoltaic systems, Prediction, Decision tree.

## 1. INTRODUCCIÓN

Hoy en día, uno de los factores externos que ha incidido al largo del tiempo en la industria es el estiaje, generando pérdidas en la producción, un gran porcentaje de industrias siguen utilizando energía convencional, siendo dependiente de las fuentes hidroeléctricas (Naciones Unidas, 2024). Considerando este problema se ha visto la necesidad de desarrollar esta investigación que permita renovar la matriz productiva de las industrias en donde se considere la migración de energía convencional a energía renovable.

El objetivo principal de esta investigación es crear una herramienta predictiva que utilice técnicas de aprendizaje automático para mejorar la eficiencia en la generación de energía fotovoltaica. El empleo del aprendizaje automático permite analizar grandes volúmenes de datos, lo cual facilita la predicción de los meses de enero, febrero y marzo del año 2024. Esto optimiza significativa los procesos energéticos, contribuyendo a la sostenibilidad y la reducción de costos operativos en el parque industrial de la ciudad de Riobamba, en la provincia de Chimborazo (Ministerio de Energía y Recursos Naturales No Renovables, 2018). Esta herramienta no solo contribuirá a maximizar la eficiencia de los paneles solares en la región, sino que también apoyará la toma de decisiones estratégicas en la gestión de recursos energéticos renovables, adaptándose a las características climáticas y geográficas específicas del área. En este estudio, se emplearon datos históricos de radiación solar obtenidos de la base de datos NASA POWER (Prediction of Worldwide Energy Resources), que están disponibles públicamente (NASA, 2024). La accesibilidad de estos datos a través de plataformas en línea facilita su uso en investigaciones que requieren información precisa y actualizada sobre recursos energéticos globales. Para la realización de este análisis, se seleccionó específicamente el recurso NASA POWER, dado que este ofrece un acceso directo y bien estructurado a datos solares, los cuales son cruciales para la evaluación de potenciales aplicaciones en energías renovables, estudios ambientales y climáticos, de los años 2017-2023, con mediciones realizadas cada hora (7:00-19:00), estos datos fueron analizados para verificar la incidencia de radiación según el horario. La radiación solar es un factor crucial que varía con el tiempo y es esencial para la generación de energía fotovoltaica. Por ello, contar con predicciones fiables y precisas de la radiación solar es fundamental para el diseño adecuado de los sistemas fotovoltaicos. Estas predicciones permiten optimizar el rendimiento y la eficiencia de los sistemas, asegurando que se dimensionen correctamente para maximizar la captación de energía solar según las condiciones específicas de cada ubicación (Solargis, 2024).

El empleo de técnicas de Aprendizaje Automático (Machine Learning) resulta crucial para el desarrollo de algoritmos que pueden prever con precisión la radiación solar. Este enfoque permite a los sistemas aprender de conjuntos de datos extensos y complejos, identificando patrones

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

y correlaciones que no son fácilmente discernibles a través de métodos tradicionales. Al adaptarse y mejorar continuamente con base en nuevos datos, el aprendizaje automático no solo aumenta la precisión de las predicciones, sino que también optimiza la eficiencia operativa de las instalaciones solares, asegurando que se maximice la captación de energía solar de acuerdo a las variaciones ambientales y climáticas específicas. Esta técnica se revela como un recurso altamente eficaz para anticipar los niveles de radiación que se presentarán en distintos momentos del día y durante diversos meses del año (Lorenzo J., 2024). Esto permite a los operadores de sistemas fotovoltaicos planificar de manera más efectiva y asegurar una optimización continua del rendimiento energético (Ordoñez-Palacios, 2020). La capacidad de realizar predicciones precisas mediante el uso de algoritmos matemáticos habilita el desarrollo de análisis detallados y la toma de decisiones estratégicas en el ámbito de la generación fotovoltaica. Esta precisión en las predicciones contribuye significativamente a la mejora de los sistemas fotovoltaicos, optimizando el uso de recursos energéticos renovables. Además, esto ayuda a disminuir la dependencia de los combustibles fósiles y promueve la utilización de fuentes de energía más limpias y libres de emisiones, alineándose con los objetivos globales de sostenibilidad y protección ambiental (Herrera A., 2023).

## 2. MATERIALES Y MÉTODOS



El desarrollo de la investigación empieza con el tratamiento de los datos, creando una base que contiene información del parámetro radiación con un tiempo de 12 horas durante el día. Como siguiente paso se identifica el porcentaje de datos que van a ser usados para el entrenamiento y prueba del modelo de árboles de decisión, las investigaciones realizados anteriormente recomiendan el uso del 80% para el inicio del proceso y el 20% para la validación del modelo. Después de configurar el modelo, se procede con la ejecución de la programación necesaria para recabar los datos predictivos de la radiación solar, empleando árboles de decisión durante el entrenamiento del modelo. Adicionalmente, se determinan las métricas de validación para evaluar la precisión y efectividad del modelo. Estos datos predictivos nos proporcionan la información esencial para efectuar los cálculos requeridos en el diseño y optimización del sistema fotovoltaico, asegurando así que se ajuste a las necesidades energéticas y las condiciones ambientales específicas.

### Normativas

Las normativas son dirigidas por el Ministerio de Energía y Recursos Naturales No Renovables a través de varios reglamentos y decretos ejecutivos. Estas regulaciones promueven la adopción de tecnologías limpias y sostenibles, reduciendo la dependencia de combustibles fósiles y disminuyendo las emisiones de carbono, la investigación tuvo como objetivo apostar al uso de fuentes renovables, siendo medidas que no solo refuerzan la responsabilidad ambiental, sino que también promueven el desarrollo económico sostenible. Las normativas vigentes son:

- Ley Orgánica del Servicio Público de Energía Eléctrica: Regula la generación, transmisión,

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

distribución y comercialización de la energía.

- Decreto Ejecutivo No. 1036: Este decreto consolidó varias agencias en la Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, centralizando la regulación del sector energético.
- Políticas del Sector Eléctrico: Establecidas por el presidente, estas políticas orientan el desarrollo y promoción de la eficiencia y sostenibilidad en el servicio de energía eléctrica.

**Escenario**

En la tabla 1 se puede observar la cantidad de datos levantados para el desarrollo de la investigación al igual que el período. Estos datos pertenecen al sector del parque Industrial de la ciudad de Riobamba ubicada geográficamente en: 8999+Q8F, Riobamba, -1.68040, -78.63170.

**Tabla 1.** Datos de la radiación solar del Parque Industrial-Riobamba

| MES | DIA | HORA | RADI-ACION 2017 | RADIA-CION 2018 | RADIA-CION 2019 | RADIA-CION 2020 | RADIA-CIÓN 2021 | RADIA-CIÓN 2022 | RADIACIÓN 2023 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 7 | 4,424 | 0,474 | 5,951 | 0,000 | 3,356 | 2,519 | 3,7785 |
| 1 | 1 | 8 | 27,858 | 19,872 | 80,967 | 0,000 | 8,455 | 7,432 | 6,345 |
| 1 | 1 | 9 | 76,638 | 28,233 | 186,836 | 0,000 | 15,634 | 19,545 | 21,344 |
| 1 | 1 | 10 | 145,765 | 49,766 | 274,132 | 21,359 | 80,324 | 90,506 | 104,0819 |
| 12 | 31 | 10 | 245,062 | 268,561 | 118,383 | 226,641 | 233,100 | 251,748 | 254,769088 |
| 12 | 31 | 11 | 327,396 | 326,627 | 223,941 | 305,993 | 314,714 | 339,891 | 343,9699958 |
| 12 | 31 | 12 | 384,868 | 341,279 | 229,828 | 270,684 | 278,399 | 300,671 | 304,278762 |
| 12 | 31 | 13 | 161,341 | 343,886 | 163,380 | 246,920 | 253,958 | 274,274 | 277,5655374 |
| 12 | 31 | 14 | 115,442 | 355,859 | 142,759 | 249,232 | 256,335 | 276,842 | 280,1637642 |
| 12 | 31 | 15 | 167,045 | 321,379 | 70,431 | 111,788 | 114,974 | 124,172 | 125,6618999 |
| 12 | 31 | 16 | 63,184 | 259,326 | 60,407 | 107,829 | 110,902 | 119,774 | 121,2115468 |
| 12 | 31 | 17 | 53,453 | 162,182 | 28,115 | 69,427 | 71,406 | 77,118 | 78,04372703 |
| 12 | 31 | 18 | 7,375 | 58,156 | 13,461 | 11,341 | 11,664 | 12,597 | 12,74832332 |
| 12 | 31 | 19 | 0,000 | 1,308 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000113076 |

*Fuente: Autores., 2024*

La investigación fue desarrollada con una cantidad de datos de 4746 de radiación solar, los mismos que fueron tomados cada hora todos los días en los años 2017-2023 de la página web de la NASA POWER. El tratamiento de los datos se realizó en el software Python eliminando las celdas con valores Nulos.

Los datos nulos pueden introducir valores erróneos en la programación, lo que resulta en errores en los resultados, afectando negativamente la precisión de las predicciones. En la figura 1 se observa cómo se eliminan los datos nulos para mejorar la calidad de la base de datos, optimizando la efectividad de la predicción.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Figura. 1**. Tratamiento de la información

```
# Eliminar filas con datos nulos
data_melted.dropna(inplace=True)
```

*Fuente: Autores, 2024*

## Selección de técnicas de Aprendizaje Automático

Existen diversos métodos y técnicas de Aprendizaje Automático para la predicción de la radiación solar. En la Tabla 2, se ilustra varias técnicas proporcionando una comparación que nos ayudará a elegir la más adecuada para el proyecto.

**Tabla 2.** Cuadro comparativo de técnica de Aprendizaje Automático

| Técnica | Ventajas | Desventajas | Predicción de radiación |
|---|---|---|---|
| Árbol de decisión | Fácil de entender e interpretar<br>No requiere normalización de datos | Pueden sobre ajustarse<br>Sensibles a pequeñas variaciones | Eficaz para detectar relaciones sencillas en datos de radiación |
| LSTM | Capacidad para aprender dependencias a largo plazo.<br>Evita problemas del desvanecimiento del gradiente. | Tiempo de entrenamiento prolongado<br>Necesita grandes volúmenes de datos | Ideal para analizar series temporales y patrones complejos de radiación. |
| RNN | Eficaz con datos secuenciales<br>Menos compleja que el LSTM | Problemas con el desvanecimiento del gradiente<br>No tiene la capacidad de aprender dependencias a largo plazo | Adecuada para los patrones de radiación de corto a medio plazo. |
| SVM | Alta precisión<br>Eficaz en espacios de alta dimensionalidad | Dificultada para interpretar parámetros<br>Requiere ajustes de parámetros | Útil para clasificar patrones de radiación |

Considerando la información presentada en la tabla 1, se puede concluir que el árbol de decisión es la técnica de aprendizaje más recomendable por varias razones. En primer lugar, su sencilla interpretación, lo cual facilita la comprensión de los resultados.

## Indexación de Datos

El proceso de indexación de datos se basa en una estructura que permite organizar y acceder a los registros de manera más eficiente. Al asignar un nuevo índice a cada registro del DataFrame, relacionado con la hora y fecha, como se observa en la figura 2, esta facilita una reestructuración de los datos. La reorganización mejora la capacidad de manejar y acceder a los datos, también optimiza el análisis y visualización de información.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Figura. 2**. Indexación de datos

```
# Combinar las columnas MES, DIA y HORA en una única columna de fecha y hora
data['FECHA_HORA'] = pd.to_datetime(data[['MES', 'DIA', 'HORA']].astype(str).agg('-'.join, axis=1), format='%m-%d-%H')

# Derretir las columnas de radiación en una sola columna para unificar los datos de diferentes años
data_melted = data.melt(id_vars=['FECHA_HORA'],
                        value_vars=['RADIACION 2017', 'RADIACION 2018', 'RADIACION 2019',
                                    'RADIACION 2020', 'RADIACIÓN 2021', 'RADIACIÓN 2022', 'RADIACIÓN 2023'],
                        var_name='AÑO',
                        value_name='RADIACION')


# Crear una nueva columna 'FECHA' que combine 'FECHA_HORA' y 'AÑO'
data_melted['FECHA'] = data_melted.apply(lambda row: row['FECHA_HORA'].replace(year=row['AÑO']), axis=1)

# Configurar la columna 'FECHA' como índice
data_melted.set_index('FECHA', drop=True, inplace=True)
```

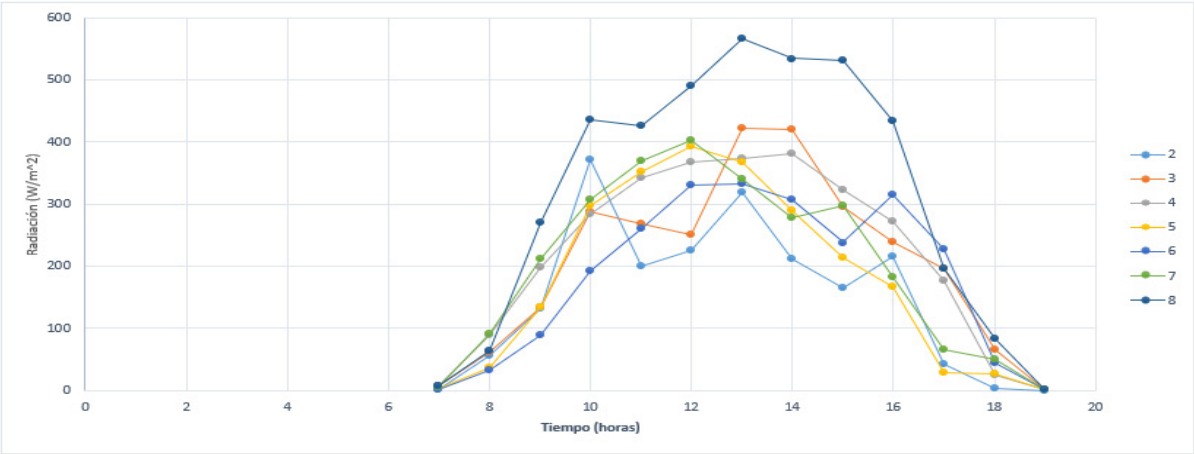
*Fuente: Autores, 2024*

## Comportamiento de la radiación solar en la zona de estudio

Analizar los datos es esencial para entender cómo varía la radiación solar a lo largo de diferentes periodos. Es particularmente importante evaluar los datos en intervalos semanales para captar las fluctuaciones en la curva de radiación solar, las cuales a menudo ocurren en picos breves. Estudiar estos patrones permite identificar tendencias relevantes que tienen un impacto directo en la eficiencia de la generación de energía solar. Reconocer estas tendencias ayuda a optimizar los sistemas fotovoltaicos y a prever comportamientos futuros de la radiación. En la figura 3 se puede observar cual es el comportamiento de la radiación solar del mes de enero de la zona en análisis.



**Figura. 3**. Comportamiento semanal de las curvas de radiación mes de enero



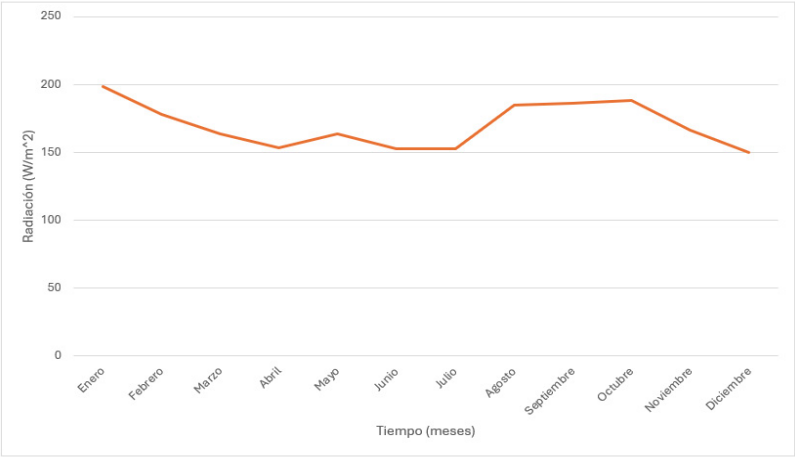

*Fuente: Autores, 2024*

El análisis de los datos promedio mensuales se enfoca en examinar las variaciones de la radiación solar a través del tiempo. Observar estos patrones es crucial para comprender la variabilidad inherente de la radiación solar y es fundamental para la optimización de la generación de energía fotovoltaica. Identificar estos patrones facilita la previsión de períodos de alta o baja producción energética, permitiendo así ajustes proactivos en la planificación energética para maximizar la eficiencia y la respuesta a la demanda. Un análisis minucioso y sostenido de los

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

datos proporciona una comprensión precisa de las variaciones de la radiación solar a lo largo del año. Esta información es vital para mejorar la eficiencia y la planificación de los sistemas fotovoltaicos, permitiendo ajustes estratégicos que optimicen el aprovechamiento de la energía solar en función de las tendencias observadas. Es crucial disponer de un análisis detallado sobre las fluctuaciones de la radiación solar mes a mes durante el año 2023. Este examen resulta fundamental para identificar los meses con mayor radiación solar, como muestra la figura 4, donde enero es señalado como el mes con la incidencia más alta. Entender estos patrones anuales de radiación facilita la proyección de la producción energética y el desarrollo de tácticas para incrementar la eficiencia, optimizando el uso de los recursos solares.

**Figura. 4**. Comportamiento de radiación promedio anual año 2023




*Fuente: Autores, 2024*

## Entrenamiento del algoritmo de árbol de decisiones

La figura 5 se muestra parte de la programación del porcentaje asignado al entrenamiento y validación del modelo. Conforme a las recomendaciones del documento (Gholamy, A., Kreinovich, V., & Kosheleva, O., 2018), sugiere una partición de datos en un 80% para el entrenamiento y un 20% para la validación en contextos de aprendizaje automático. Esta proporción garantiza que el modelo se ajuste adecuadamente a los datos de entrenamiento. Además, utilizar un 20% de los datos para la validación permite probar la capacidad predictiva del modelo, evitando el riesgo de sobreajuste y asegurando así la fiabilidad de las predicciones (Wang, F., Mi, Z., Su, S., & Zhao, H., 2012). Se utilizan datos del período 2017-2022 y una parte del año 2023 como datos de entrenamiento. El resto del año 2023 se emplea como datos de validación.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Figura. 5**. Porcentaje de datos para entrenamiento de modelo

```
# Dividir los datos en conjuntos de entrenamiento y prueba (80% y 20%)
split_index = int(len(X) * 0.8)
X_train, X_test = X[:split_index], X[split_index:]
y_train, y_test = y[:split_index], y[split_index:]
```

*Fuente: Autores, 2024*

## Validación del algoritmo de árbol de decisiones

Hay diversas técnicas para validar modelos predictivos, incluida la matriz de confusión, que se ilustra en la tabla 3. Esta tabla categoriza los resultados en verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Las entradas en la diagonal principal de la matriz corresponden a clasificaciones acertadas, mientras que las entradas fuera de esta diagonal reflejan errores en la clasificación. Este método es esencial para evaluar la precisión y efectividad de un modelo de predicción.

**Tabla 3.** Valores de la matriz de confusión

| | **Predicho Negativo** | **Predicho Positivo** |
|---|---|---|
| Real Negativo | TN | FP |
| Real Positivo | FN | TP |

## Precisión

La métrica conocida como precisión se calcula como la proporción de verdaderos positivos en relación al total de predicciones positivas, tal como se señala en (James, G., Witten, D., Hastie, T., & Tibshirani, R., 2017). Esto significa que la precisión indica cuántas de las instancias que el modelo clasificó como positivas son efectivamente positivas, tal como se formula en la ecuación 1. Es crucial considerar que una precisión muy alta podría sugerir que el modelo está sobreajustado, lo cual implica que, aunque realiza predicciones precisas sobre los datos de entrenamiento, podría no generalizar bien con nuevos datos.

$$Precisión = \frac{TP}{TP + FP} \quad (1)$$

## Sensibilidad

La sensibilidad, o tasa de verdaderos positivos, se define como la proporción de verdaderos positivos dividida por el total de casos positivos reales. Esta métrica, es esencial para evaluar la capacidad del modelo de identificar todas las instancias positivas de manera correcta. Su formulación matemática está detallada en la ecuación 2. Al medir la sensibilidad, podemos comprender cuán efectivo es el modelo al no pasar por alto los verdaderos casos positivos, lo que es particularmente crucial en contextos donde omitir un positivo tiene consecuencias significativas.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

$$Sensibilidad = \frac{TP}{TP + FP} \quad (2)$$

**Exactitud**

La precisión global del modelo se determina calculando la proporción de predicciones correctas, que incluye tanto los verdaderos positivos como los verdaderos negativos, respecto al total de casos evaluados. Esta medida, se visualiza en la ecuación 3, donde se presenta su formulación matemática. La precisión global es fundamental para valorar la efectividad general del modelo en clasificar correctamente todas las instancias, ya sean positivas o negativas, proporcionando una visión integral de su rendimiento.

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

En la figura 6 se muestra un extracto de la programación para validar el modelo de árbol de decisión en el software Python.



**Figura. 6.** Métricas para validación del modelo

```
y_pred_bin = (etiquetas_predichas[ RADIACION_PRED ] > 0

# Calcular la matriz de confusión
cm = confusion_matrix(y_test_bin, y_pred_bin)

# Extraer valores de la matriz de confusión
TN, FP, FN, TP = cm.ravel()

# Calcular precisión, sensibilidad y exactitud
precision = TP / (TP + FP)
sensibilidad = TP / (TP + FN)
exactitud = (TP + TN) / (TP + TN + FP + FN)
```

*Fuente: Autores, 2024*

**Predicción de la radiación solar**

Para la implementación del algoritmo de árbol de decisiones es importante conocer la composición del modelo.

**Algoritmo de árbol de decisiones**

El modelo de árbol de decisiones utiliza criterios específicos de división para evaluar la calidad de las separaciones en los conjuntos de datos, como se indica en (Wang, 2012). Las ecuaciones 4, 5 y 6 ilustran estas reglas matemáticas, que son cruciales para la construcción efectiva del modelo. Estos criterios permiten al modelo segmentar los datos de manera que cada división mejore la precisión de las predicciones al crear subconjuntos más homogéneos, fundamentales

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

para la operatividad y eficacia del árbol de decisiones.

**Impureza de Gini:**

$$Gini(D) = \sum_{i=1}^{k} pi^2 \quad (4)$$

Donde:

$pi$ *es la muestra de ejemplos en la clase i en el nodo D*

$k$ *número de clases*

**Varianza**

$$Varianza = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (5)$$

Donde:

$N$ *es el número de instancias*

$y_i$ *valor real de cada instancia*

$\bar{y}$ *valor promedio de las instancias en el nodo*

**Función objetivo**

$$\hat{y} = \frac{1}{K}\sum_{i=1}^{K} y_i \quad (6)$$

Donde:

$\hat{y}$ *predicción para la nueva instancia que llega a un nodo hoja especifico.*

$K$ *número de instancias que llegaron a ese nodo hoja*

$y_i$ *valor real en cada una de las instancias en el nodo hoja*

**Métricas de evaluacion del modelo**

En la figura 7 se muestra la programación de las métricas de evaluación del modelo.

Error cuadrático medio (MSE): es el cálculo de la media de los cuadrados de los errores, presentado en la ecuación 7.

$$MSE = \frac{1}{N}\sum_{i==1}^{N}(y_i - \hat{y})^2 \quad (7)$$

Coeficiente de determinación ($R^2$): Indicador estadístico empleado para valorar la adecuación del modelo se muestra en la ecuación 8.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y})^2}{\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad (8)$$

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Figura. 7**. Calculo de los indicadores MSE, NME,

```
# Calcular métricas de error
mse = mean_squared_error(etiquetas_reales['RADIACION'], etiquetas_predichas['RADIACION_PRED'])
nme = mse / np.mean(etiquetas_reales['RADIACION'])
r2 = r2_score(etiquetas_reales['RADIACION'], etiquetas_predichas['RADIACION_PRED'])
```

*Fuente: Autores., 2024*

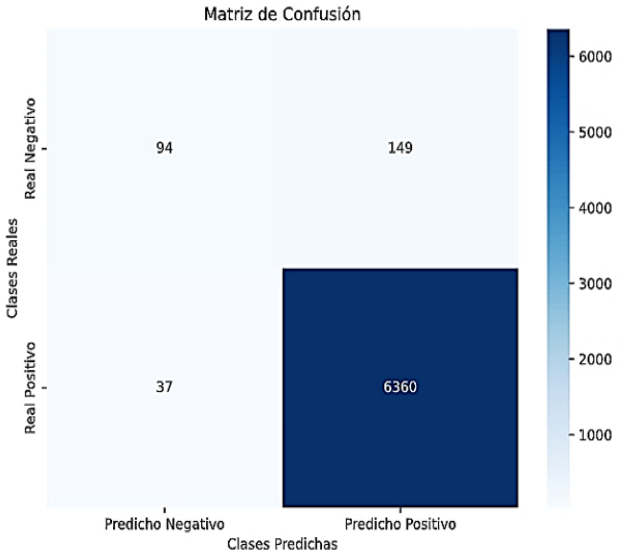
## 3. RESULTADOS Y DISCUSIÓN

A continuación, se detallan los resultados del proyecto de investigación, que abarcan tanto la fase de predicción y validación del modelo como los datos específicos de predicción para un periodo de tres meses. Además, se incluyen los resultados obtenidos del desempeño del sistema fotovoltaico durante este tiempo. Esta información es fundamental para evaluar la precisión y la efectividad del modelo en condiciones reales, proporcionando una base sólida para futuras mejoras y ajustes en el sistema.

### Matriz de confusión

La matriz de confusión calculada se presenta en la figura 8, donde se observan 6360 verdaderos positivos (TP), 149 falsos positivos (FP), 94 verdaderos negativos (TN), y 37 falsos negativos (FN). Estos resultados muestran una predominancia de verdaderos positivos sobre los falsos negativos, indicativo de que el modelo posee una excelente capacidad para identificar correctamente los casos positivos. Esto se refleja en un alto nivel de sensibilidad, lo que demuestra que el modelo es confiable y efectivo en la detección de eventos o condiciones positivas, justificando su uso para realizar predicciones fiables.

**Figura. 8**. Matriz de confusión par la validación del entrenamiento




*Fuente: Autores., 2024*

Además, se consideran otras métricas de validación del modelo, como la precisión, exactitud y



Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático. pp. 338 - 355 / Volumen 6, número 1 / DOI: https://doi.org/10.37431/conectividad.v6i1.196

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

sensibilidad. En la tabla 4 se presentan los valores de estas métricas.

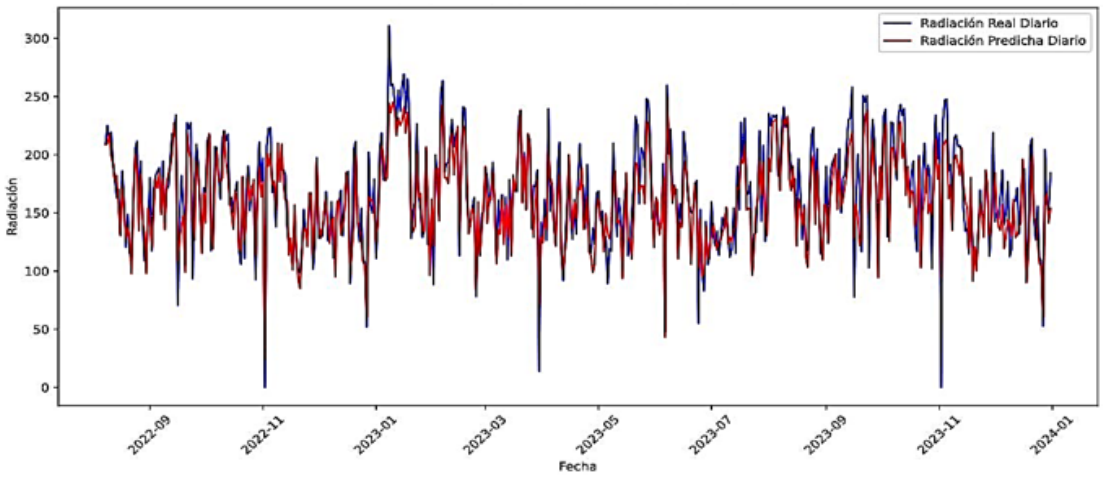
**Tabla 4.** Resultados de las métricas de evaluación

| Parámetro | Valor |
|---|---|
| Precisión | 0.9 |
| Sensibilidad | 0.9 |
| Exactitud | 0.9 |
| MSE | 4267.3 |
| NME | 21,3 |
| Coeficiente de determinación | 0.70 |

## Valores de prueba y predicción

Es crucial adoptar un enfoque que permita la predicción diaria de la radiación solar, como se ilustra en la figura 9. Utilizando el 80% de los datos para entrenar el modelo y el 20% restante para su validación, es posible verificar la precisión del modelo. En este contexto, la comparación entre el comportamiento real y los datos predichos muestra una gran similitud, lo que indica que el modelo ha captado de manera efectiva las tendencias y patrones diarios de la radiación solar. Esto subraya la fiabilidad del modelo para predecir la radiación solar de forma precisa en un marco diario.

**Figura. 9**. Comparación de valores de prueba y valores de predicción diaria.




*Fuente: Autores, 2024*

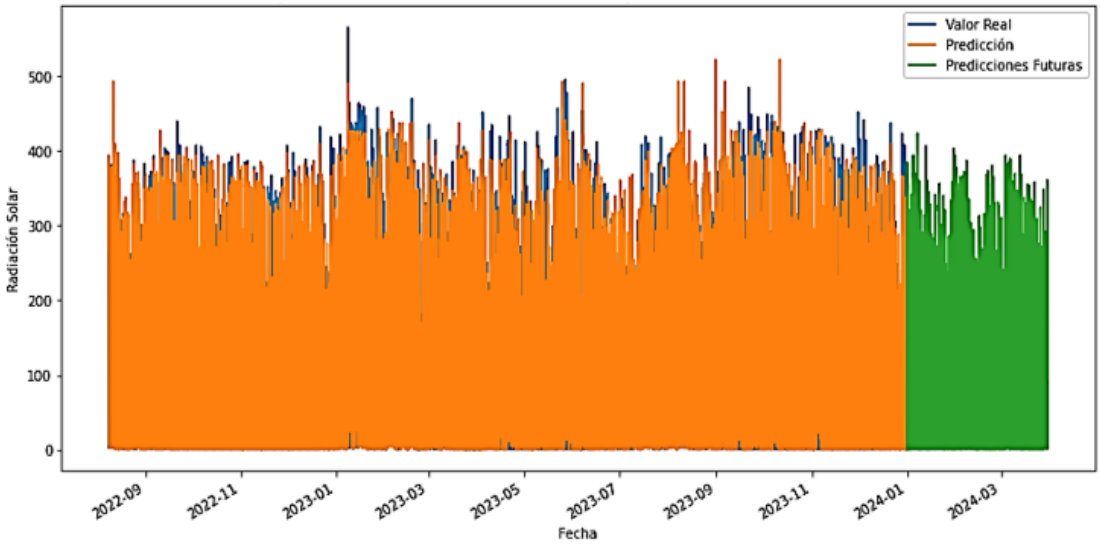
## Predicción del mes de enero, febrero y marzo del 2024

El análisis mostrado en la figura 10 compara los datos actuales de radiación solar con las predicciones futuras de los meses de febrero y marzo. Inicialmente, estas predicciones se validan al compararlas con los datos reales, lo que demuestra una buena precisión del modelo debido a la similitud en el comportamiento observado. Posteriormente, se lleva a cabo una predicción para los próximos tres meses, utilizando la desviación estándar como indicador clave para evaluar la

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

incertidumbre asociada con estas predicciones futuras. Esta metodología es fundamental para confirmar la efectividad del modelo en prever cambios y tendencias en la radiación solar con un margen de confianza razonable.

**Figura. 10**. Comparación de valor real, validación y predicciones




*Fuente: Autores, 2024*

## Dimensionamiento del sistema fotovoltaico



Para dimensionar adecuadamente el sistema, se llevaron a cabo estudios específicos en la zona en cuestión. Se tomaron en cuenta parámetros de diversas fuentes, incluyendo geoportales y otros sitios web relevantes. Por ejemplo, la cantidad de luminarias se estimó basándose en comparaciones con lugares que presentan características similares a la zona de estudio. La eficiencia de los paneles solares se determinó de acuerdo al tipo de panel elegido, y los días de autonomía se calcularon según las necesidades específicas del cliente. Utilizando esta información se configuraron los datos de entrada que aparecen en la tabla 5 para realizar el análisis pertinente con un usuario de tipo comercial.

**Tabla 5.** Variables de entrada para el dimensionamiento del sistema fotovoltaico

| Variable | Valor |
|---|---|
| Consumo mensual total en kWh | 3500 |
| Número de clientes | 5 |
| Potencia de las luminarias en W | 600 |
| Número de luminarias | 23 |
| Horas de funcionamiento diarias de las luminarias | 8 |
| Eficiencia del panel solar | 0.8 |
| Voltaje del sistema (kV) | 13.8 |
| Factor de sobrecarga del inversor | 1.25 |
| Margen de seguridad | 0.2 |
| Potencia nominal de cada panel solar en W | 300 |
| Eficiencia de la batería | 0.9 |

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

| Variable | Valor |
|---|---|
| Capacidad máxima de la batería en kWh | 200 |
| Energía inicial almacenada en la batería en kWh | 20 |

En la Tabla 6 se presentan los datos calculados para el sistema fotovoltaico.

**Tabla 6.** Cálculo del sistema fotovoltaico

| | |
|---|---|
| Consumo mensual promedio por cliente | 700 kWh |
| Consumo diario promedio por cliente | 23.33 kWh |
| Energía diaria requerida por las luminarias | 120.00 kWh |
| Energía diaria total requerida | 123.5 kWh |
| Energía necesaria para los paneles solares | 0.89 kWh |
| Área requerida para los paneles solares | 4.95 m² |
| Energía total necesaria para la batería | 123.5 kWh |

**Energía diaria total requerida**

Se hace una estimación de consumo de los siguientes elementos considerando que el parque industrial de Riobamba es una zona industrial:

1. **Prensa**,
   - **Marca Ejemplo:** Trumpf, Amada
   - **Consumo:** 5 kWh por hora
   - **Estimación (50 kWh)**
2. **Tornos CNC**
   - **Marca Ejemplo:** Mazak, Haas
   - **Consumo:** 3 kWh por hora
   - **Estimación (30 kWh)**
3. **Máquinas de Moldeo por Inyección**
   - **Marca Ejemplo**: Engel, Husky
   - **Consumo**: 2 kWh por hora
   - **Estimación** (20 kWh)
   - **3 Televisores** = 3 * 1.5 kWh = 4.5 kWh
   - **5 Computadoras** = 4 * 1.5 kWh = 6 kWh
   - **Luminarias =** 13 kWh

Sumando estos consumos tendríamos: 123.5kWh

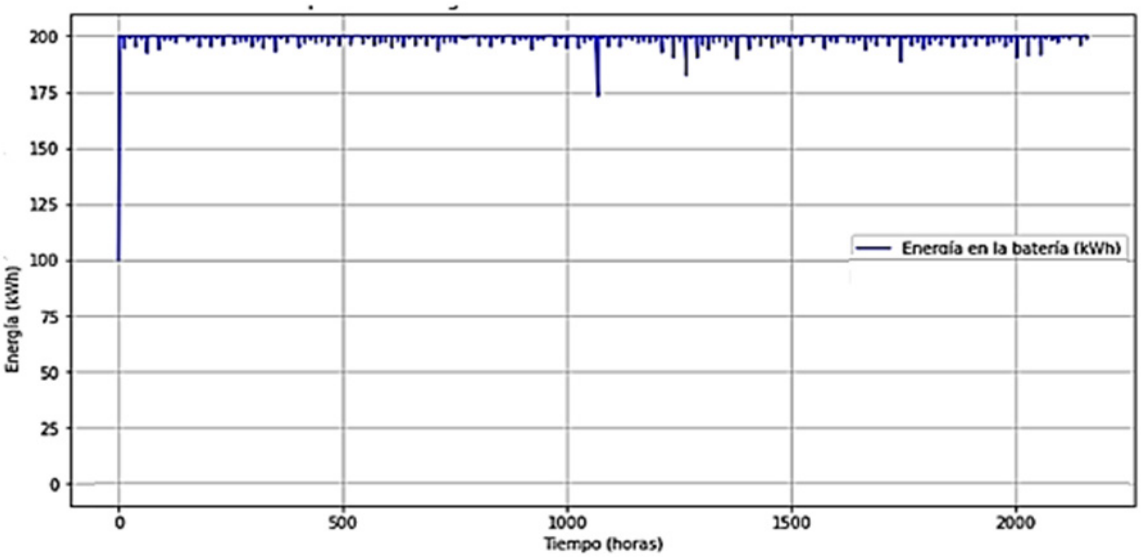
**Energía almacenada en la batería**

La figura 11 ofrece una comparación detallada entre la energía almacenada en la batería. Se observa que la cantidad de energía almacenada es significativamente alta, lo cual indica que el sistema de almacenamiento está operando eficientemente y conserva una reserva sustancial de

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

energía. Esto queda evidenciado por la línea azul del gráfico, que muestra muy pocos picos de descarga, reflejando la capacidad del sistema para gestionar el consumo sin comprometer las reservas energéticas.

**Figura. 11**. Energía almacenada en la batería

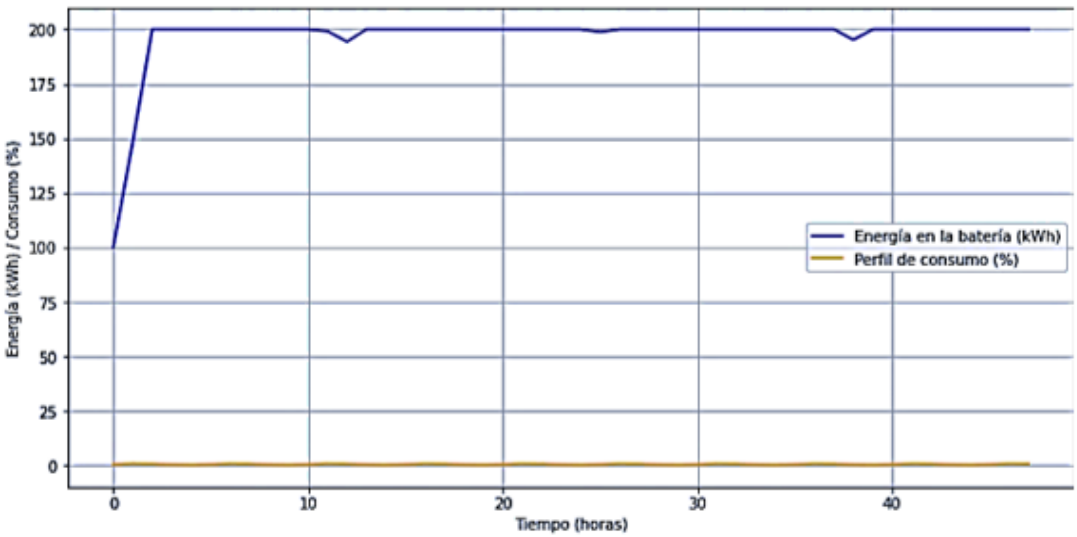



*Fuente: Autores, 2024*



La Figura 12 ilustra cómo el dimensionamiento del sistema permite que la batería opere cerca de su capacidad máxima, mientras que el consumo de energía se mantiene relativamente bajo, teniendo en cuenta que el perfil de consumo se ilustra como porcentaje en comparación con la cantidad almacenada que es en kWh. Este comportamiento indica que el sistema está diseñado para proporcionar un suministro de energía constante sin el riesgo de agotar las reservas de la batería, asegurando así la fiabilidad y la eficiencia en la gestión de la energía almacenada.

**Figura. 12**. Comparación de la batería y el perfil de consumo.



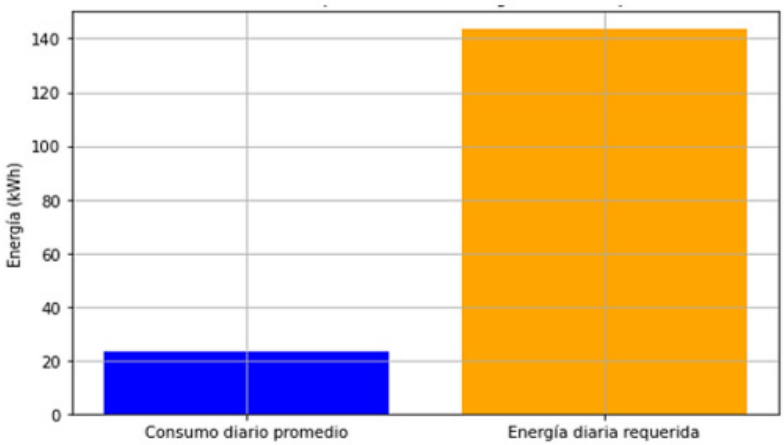

*Fuente: Autores, 2024*

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

## Consumo promedio en comparación de la energía diaria requerida

En la Figura 13 se presenta una gráfica que muestra que el sistema fotovoltaico está dimensionado para cubrir una demanda de energía considerablemente mayor que el consumo promedio diario. Esta capacidad adicional es beneficiosa ya que permite al sistema satisfacer otras necesidades eléctricas y además los días con menos radiación solar, el sistema puede seguir operando de manera eficiente son limitaciones.

**Figura. 13**. Consumo diario promedio vs energía diaria requerida

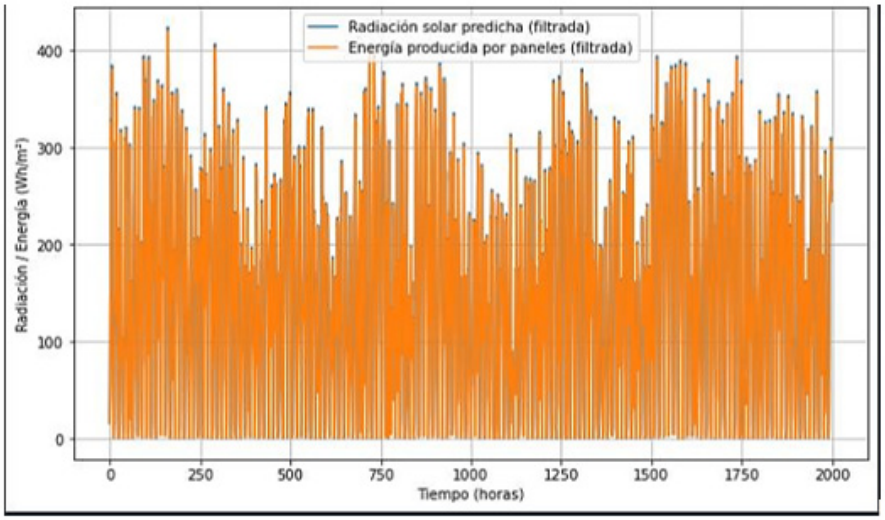



*Fuente: Autores, 2024*

La Figura 14 ilustra la relación entre la radiación solar predicha y la energía generada por los paneles solares a lo largo de 2000 horas. Se observan picos de radiación que alcanzan más de 400 en varios intervalos, señalando períodos de intensa radiación solar. Existe una correspondencia evidente entre la radiación solar pronosticada y la energía producida por los paneles, lo cual coincide con el desempeño esperado de un sistema fotovoltaico eficaz. La energía generada refleja adecuadamente las variaciones en la radiación disponible, demostrando la capacidad del sistema para convertir la energía solar en eléctrica de manera eficiente.

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

**Figura. 14**. Radiación solar predicha vs Energía producida por los paneles (200 horas).




*Fuente: Autores, 2024*

## 4. CONCLUSIONES

El coeficiente de determinación ($R^2$) fue de 0.70 siendo la métrica más importante. Este valor indica que el modelo explica el 70% de la variabilidad en la radiación solar, lo cual es significativo para asegurar predicciones precisas. Esta precisión en las predicciones resulta en una mejor planificación y optimización del uso de los paneles solares, maximizando así la eficiencia de la generación de energía en el contexto del Parque Industrial de Riobamba.

El uso de diversas métricas, incluyendo precisión, exactitud, sensibilidad, matriz de confusión y error cuadrático medio, es fundamental para evaluar la eficiencia y el rendimiento de los modelos de predicción. Cada una de estas métricas ofrece una vista única sobre la calidad del modelo. Al combinarlas, se facilita una evaluación exhaustiva del modelo, permitiendo identificar su capacidad predictiva de manera más efectiva. Esta evaluación integral es esencial para comprender cómo cada modelo maneja diferentes tipos de datos y situaciones, asegurando que las decisiones basadas en estos modelos sean lo más precisas posible.

El diseño del sistema fotovoltaico para el área del parque industrial se ha basado en un minucioso análisis de la radiación solar y los requisitos energéticos específicos de los usuarios comerciales. El sistema propuesto, equipado con una capacidad pico de 1,26 kWp y un esquema de almacenamiento óptimo, está diseñado para cubrir una demanda energética diaria de 123,5 kWh, que incluye tanto el consumo de los clientes como la operación de las luminarias. Con una capacidad utilizable de la batería de 286,77 kWh, el sistema asegura un suministro constante de energía, incluso en períodos de baja radiación solar. Además, la potencia del inversor se ha calculado cuidadosamente para gestionar las cargas de manera eficiente y proporcionar un margen de seguridad adecuado. Esta planificación garantiza que el sistema no solo cumpla con

Fecha de recepción: 29 / 09 / 2024
Fecha de aceptación: 06 / 12 / 2024
Fecha de publicación: 23 / 01 / 2025

las necesidades actuales, sino que también ofrezca fiabilidad y eficiencia en su funcionamiento.

## REFERENCIAS

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Recuperado de* https://scholarworks.utep.edu/cs_techrep/1209

Herrera Jiménez, A. (2023). Análisis y predicción de radiación en sistemas fotovoltaicos haciendo uso de Aprendizaje Automático .

Lorenzo J. A. (2024)*, "Radiación, Irradiación y Azimut en Fotovoltaica*

. *SunFields."* [Online]. Available: https://www.sfe-solar.com/noticias/articulos/energia-

fotovoltaica-radiacion-geometria-recorrido-optico-irradiancia-y-hsp/#Irradiancia

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning with applications in R*. Springer Science+Business Media.

Ministerio de Energía y Recursos Naturales No Renovables. (2018). Plan Maestro de Electricidad (pp. 54–55). Recuperado el 8 de julio de 2024, de https://www.celec.gob.ec/wp-content/uploads/2023/02/Plan-Maestro-de-Electricidad.pdf

NASA. (2024). NASA POWER | Predicción de los recursos energéticos mundiales. *Recuperado el 17 de julio de 2024, de* https://power.larc.nasa.gov/

Ordoñez-Palacios, L.-E., et al. (2020). Predicción de radiación solar en sistemas fotovoltaicos utilizando técnicas de aprendizaje automático. *Revista Facultad de Ingeniería, 29(54), e11751.* https://doi.org/10.19053/01211129.V29.N54.2020.11751

Solargis. (2024). Previsiones solares y predicción solar - Visión general. *Recuperado el 8 de agosto de 2024, de* https://solargis.com/es/products/forecast

Wang, F., Mi, Z., Su, S., & Zhao, H. (2012). Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies (Basel)*, 5(5), 1355-1370. https://doi.org/10.3390/EN5051355.

United Nations. (2024*).* La promesa de la energía solar: Estrategia energética para reducir las emisiones de carbono en el siglo XXI. Naciones Unidas. *Recuperado el 8 de agosto de 2024, de* https://www.un.org/es/chronicle/article/la-promesa-de-la-energia-solar-estrategia-energetica-para-reducir-las-emisiones-de-carbono-en-el