

Estimation of Global Solar Radiation from Extreme Temperatures Using Machine Learning Techniques

Ciro William Taipe Huaman¹; Eva Genoveva Mendoza Mamani²; Matias Huillca Arbieto³; Hugo Hernan Flores Laime⁴,

¹Universidad Nacional de Juliaca, Perú, c.taipe@unaj.edu.pe

^{2,3}Universidad Nacional del Altiplano, Perú, emendoza@unap.edu.pe, mhuillca@unap.edu.pe

⁴Universidad Nacional Intercultural de Quillabamba, Perú, hugo.flores@uniq.edu.pe

Abstract– This study aimed to model and predict the daily maximum solar radiation in the city of Puno (Peru) using meteorological variables associated with extreme temperatures through machine learning techniques. Records obtained from a DAVIS Vantage Pro 2.0 meteorological station were used, covering the period 2017–2024 and comprising 2,060 daytime observations. Three tree-based models were developed: Random Forest, XGBoost, and LightGBM, whose performance was evaluated using error and goodness-of-fit metrics, including root mean square error, mean absolute error, and the coefficient of determination. Exploratory analysis revealed a strong positive correlation between observed solar radiation and extraterrestrial radiation, confirming its role as the main physical forcing of the system. Among the evaluated models, Random Forest achieved the best predictive performance, with a root mean square error of 124.24 W/m², a mean absolute error of 96.50 W/m², and a coefficient of determination of 0.5135, outperforming XGBoost and LightGBM, which obtained coefficients of determination of 0.4482 and 0.3679, respectively. Variable importance analysis consistently identified extraterrestrial radiation and daily thermal amplitude as the most influential predictors. Overall, the results demonstrate that artificial intelligence approaches constitute effective and reliable tools for local solar radiation estimation in high-altitude regions, with potential applications in energy planning, environmental management, and climate studies.

Keywords– Extreme temperatures, artificial intelligence, solar radiation, statistical analysis

Estimación de la radiación solar global basada en temperaturas extremas mediante técnicas de inteligencia artificial

Ciro William Taipe Huaman¹; Eva Genoveva Mendoza Mamani²; Matias Huillca Arbieto³; Hugo Hernan Flores Laime⁴,

¹Universidad Nacional de Juliaca, Perú, c.taipe@unaj.edu.pe

^{2,3}Universidad Nacional del Altiplano, Perú, emendoza@unap.edu.pe, mhuillca@unap.edu.pe

⁴Universidad Nacional Intercultural de Quillabamba, Perú, hugo.flores@uniq.edu.pe

Resumen— Este estudio tuvo como objetivo modelar y predecir la radiación solar máxima diaria en la ciudad de Puno (Perú) utilizando variables meteorológicas asociadas a temperaturas extremas, mediante técnicas de aprendizaje automático. Se emplearon registros obtenidos de una estación meteorológica DAVIS Vantage Pro 2.0 correspondientes al periodo 2017–2024, que comprenden 2060 observaciones diurnas. Se desarrollaron tres modelos basados en árboles de decisión: Random Forest, XGBoost y LightGBM, cuyo desempeño fue evaluado mediante indicadores de error y ajuste, como el error cuadrático medio, el error absoluto medio y el coeficiente de determinación. El análisis exploratorio reveló una fuerte correlación positiva entre la radiación solar observada y la radiación extraterrestre, confirmando su rol como principal forzamiento físico del sistema. El modelo Random Forest presentó el mejor rendimiento predictivo, con un error cuadrático medio de 124.24 W/m², un error absoluto medio de 96.50 W/m² y un coeficiente de determinación de 0.5135, superando a XGBoost y LightGBM, que alcanzaron coeficientes de determinación de 0.4482 y 0.3679, respectivamente. El análisis de importancia de variables identificó consistentemente a la radiación extraterrestre y a la amplitud térmica diaria como los predictores más influyentes. En conjunto, los resultados evidencian que los enfoques de inteligencia artificial constituyen herramientas eficaces y confiables para la estimación local de la radiación solar en regiones de alta montaña, con aplicaciones en la planificación energética, la gestión ambiental y los estudios climáticos.

Palabras clave— Temperatura extremas, inteligencia artificial, radiación solar, análisis estadístico.

I. INTRODUCCIÓN

La demanda energética se incrementa de manera acelerada como consecuencia del avance tecnológico continuo. Para satisfacer la demanda de energía al ritmo actual, los combustibles fósiles por sí solos no pueden ser fiables debido a sus limitadas reservas y al impacto perjudicial que tienen en el medio ambiente [1].

Los datos de las mediciones detalladas de la radiación solar son de importancia para muchas aplicaciones basadas en la energía solar [2]. Si bien es posible instalar instrumentación especializada para medir con precisión la radiación solar, esta no se encuentra disponible en numerosas zonas del país.

La incidencia de la radiación solar depende de la ubicación geográfica y de las condiciones geométricas del sitio; por ello, resulta necesario desarrollar modelos que

permitan estimar la radiación solar global diaria, especialmente en zonas con alto potencial solar, a fin de optimizar el aprovechamiento de este recurso energético.

En la actualidad, la información de los valores de la radiación solar se obtiene de las bases de datos proporcionadas en la web, como la que ofrece SENAMHI, que son valores promedio diarios para grandes regiones, que difieren de los valores estimados en tiempo real, debido a las características del lugar y el tipo de clima que se presenta [3]. Debido a ello los valores promedio diarios de la radiación solar (UVI) difieren de las estimadas en tiempo real, considerando que la intensidad de la radiación solar varía durante el día y con un periodo anual [4].

Existen numerosos modelos de estimación de radiación solar basados en humedad relativa [5]–[7], en la cobertura de nubes en el cielo [5], [8], [9], basados en la posición del sol [5], [10], [11], de las temperaturas extremas [1], [5], [12], redes neuronales artificiales [13]–[15], dependientes de ajuste estadísticos [16]–[18] y procesamiento de imágenes satelitales [19], [20], en este trabajo analizamos modelos de inteligencia artificial para estimar la radiación solar global en función a las temperaturas extremas diarias.

La llegada de sistemas predictivos expertos inspirados en IA puede abordar con solidez las desventajas de los métodos mecanicistas y deterministas, en particular en la predicción de la radiación solar. Las plataformas de predicción ML y DL basadas en IA son conocidas por exhibir una inmensa eficiencia computacional y estos algoritmos son muy hábiles en el manejo de la no linealidad en los conjuntos de datos de entrada [21]. El potencial de la técnica ML ha sido explorado en la predicción de la radiación solar a largo plazo a escala global utilizando una red neuronal artificial (RNA) multicapa de propagación hacia adelante [22]. Los avances en inteligencia artificial (IA) ofrecen una alternativa robusta. Modelos híbridos y explicables como el EJH-X-DNN [23] y el W-O-convLSTM [24] han mostrado alta precisión en predicciones del UVI, alcanzando correlaciones superiores a 0,96. Asimismo, la IA permite integrar datos satelitales, meteorológicos y geográficos con técnicas de optimización que mejoran la exactitud y ofrecen interpretabilidad [25]–[27].

El empleo de técnicas de inteligencia artificial (IA) en la estimación de la radiación solar y ultravioleta ha mostrado avances significativos en los últimos años. Diversos estudios se han enfocado en modelos híbridos que combinan arquitecturas profundas y métodos de aprendizaje automático para mejorar la precisión frente a la variabilidad atmosférica. En Australia, Ghimire, Deo, et al. (2022) desarrollaron el modelo híbrido CXGBRFR, que integra CNN, Gradient Boosting, Random Forest y optimización Harris Hawks, alcanzando un desempeño superior frente a modelos convencionales en la predicción de radiación solar global [21]. De manera similar, Ghimire, Bhandari, et al., (2022) propusieron el modelo CSVN, basado en CNN y SVR, optimizado con selección de características, mostrando resultados más precisos que ocho métodos de referencia [28]. En el ámbito europeo, Guijo-Rubio et al. (2020) lograron predicciones altamente precisas de radiación solar en España utilizando redes neuronales evolutivas entrenadas exclusivamente con datos satelitales [29]. En la misma línea, Peng et al. (2021) desarrollaron un modelo híbrido que combina descomposición CEEMDAN, BiLSTM y optimización mediante algoritmo seno-coseno (SCA), superando a siete técnicas de referencia en predicciones multihorizonte [30].

Las revisiones recientes refuerzan la pertinencia de los modelos híbridos. Guermoui et al. (2020) concluyen que los enfoques individuales son insuficientes para capturar la variabilidad atmosférica, destacando que la integración de múltiples técnicas de IA mejora la precisión y la aplicabilidad en contextos solares [31]. Más recientemente, Naveed et al. (2024) introdujeron el modelo TIR, que combina BiLSTM, GRU y mecanismos de atención, alcanzando un desempeño superior ($R^2 = 0,9983$), evidenciando el potencial de arquitecturas avanzadas en el modelado de variables atmosféricas [27].

En conjunto, estos avances evidencian que el uso de modelos y arquitecturas profundas basadas en IA constituye una estrategia prometedora para la estimación de la radiación solar, al integrar de manera eficaz datos atmosféricos, satelitales y meteorológicos para aplicaciones en salud, energía y gestión ambiental.

Por lo tanto, la presente investigación se justifica en la necesidad de desarrollar un modelo predictivo basado en IA que permita estimar valores de radiación solar en función a temperaturas extremas de forma localizada en regiones de alta radiación como Puno.

II. MATERIALES Y MÉTODOS

La presente investigación se enmarca en un enfoque cuantitativo, de tipo aplicado, con diseño no experimental y alcance predictivo. El estudio es de carácter observacional y retrospectivo, basado en el análisis de series temporales de datos meteorológicos, con el objetivo de modelar la radiación

solar global a partir de variables térmicas mediante técnicas de inteligencia artificial.

A. Lugar de estudio

La investigación se desarrolló en la Universidad Nacional del Altiplano (UNAP), ubicada en la ciudad de Puno, al sur del Perú. El área de estudio se localiza en las coordenadas geográficas 15.83° de latitud sur y 70.083° de longitud oeste, a una altitud aproximada de 3 820 m s. n. m.

B. Datos atmosféricos en la ciudad de Puno

Se emplean registros del 2017-2024 obtenidos a través de la estación meteorológica DAVIS Vantage PRO 2.0 Plus [32]. En intervalos de 15min considerando datos de en el intervalo de 6:00am 18:00pm. El dispositivo dispone de certificación de calibración emitida por el proveedor, lo que garantiza la confiabilidad de las mediciones. Como se muestra en la Figura 1, la estación meteorológica está instalada en un área despejada del pabellón de Ingeniería Civil de la UNAP, libre de sombras y obstáculos que puedan interferir en el registro de datos. Su ubicación corresponde a las coordenadas -15.4890 de latitud, -70.1515 de longitud y una altitud de 3 825 m s. n. m. Para el presente estudio se consideraron las variables medidas de radiación solar y temperaturas extremas diarios.



Fig. 1 Estación meteorológica de la UNAP.

C. Procesamiento de datos

En la construcción del modelo propuesto para pronosticar el UVI en el horizonte horario, la primera etapa consiste en la extracción de datos. De la estación meteorológica para la ciudad de Puno. En el periodo 01 de enero del 2017 al 31 de diciembre del 2024, considerando datos medidos de diurno.

La segunda etapa del modelado radiación solar. Fue importante examinar cuidadosamente los conjuntos de datos extraídos para localizar cualquier dato faltante. Los datos se

limpiaron eliminando registros nocturnos. La imputación de ausentes se realizó por interpolación temporal.

En la tercera etapa, se llevó a cabo un Análisis Exploratorio de Datos (EDA) empleando diferentes técnicas estadísticas y gráficas. En primer lugar, se generaron histogramas para analizar la distribución de las variables y detectar posibles sesgos o asimetrías. Posteriormente, se construyó una matriz de correlación con el fin de identificar relaciones lineales entre las variables. Además, se aplicaron los gráficos de Autocorrelación (ACF) y Autocorrelación Parcial (PACF), útiles para evaluar la dependencia temporal de las series.

La cuarta etapa Se dividieron los datos manteniendo la secuencia temporal: 80% para entrenamiento y 20% para prueba. Para la validación e hiperparametrización se empleó TimeSeriesSplit (k=4)".

En la quinta etapa, se diseña el modelo. Para desarrollar la secuencia de modelos, se implementó un lenguaje de programación Python de renombre a través de la plataforma Google Colab con una interfaz Jupyter Notebook de libre acceso, compatible con una unidad de procesamiento tensorial (TPU) y una unidad de procesamiento gráfico (GPU). El entorno virtual de la potente herramienta Python proporciona paquetes excepcionales para la ejecución de algoritmos de aprendizaje automático (ML) y aprendizaje automático (DL).

D. Estadísticos de evaluación

La capacidad predictiva y la concordancia entre los valores observados y estimados por los modelos fueron cuantificadas mediante un conjunto de métricas estadísticas de desempeño ampliamente utilizadas en la validación de modelos meteorológicos y energéticos. Se emplearon el error cuadrático medio (RMSE), el error absoluto medio (MAE), el coeficiente de determinación (R^2) y el coeficiente de correlación lineal de Pearson (r).

El RMSE y el MAE permiten evaluar la magnitud promedio de las discrepancias entre las mediciones experimentales y las predicciones del modelo, proporcionando medidas de exactitud global expresadas en las mismas unidades de la variable analizada. Por su parte, el coeficiente R^2 cuantifica la proporción de la variabilidad observada explicada por el modelo, mientras que el coeficiente r determina la intensidad y dirección de la relación lineal entre ambas series de datos [33]. Están dados por las siguientes ecuaciones:

$$RMSE = \left[\frac{\sum_{i=1}^n (R_{mi} - R_{si})^2}{n} \right]^{1/2} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^n |R_{mi} - R_{si}|}{n} \quad (2)$$

$$R^2 = \frac{\left[\sum_{i=1}^n (R_{mi} - \bar{R}_{mi})(R_{si} - \bar{R}_{si}) \right]^2}{\left[\sum_{i=1}^n (R_{mi} - \bar{R}_{mi})^2 \right] \left[\sum_{i=1}^n (R_{si} - \bar{R}_{si})^2 \right]} \quad (3)$$

$$r = \sqrt{R^2} \quad (4)$$

Donde: R_{mi} es el dato medido por la estación meteorológica, R_{si} es el dato obtenido por modelo, $\bar{R}_{mi}$ es el promedio de los datos medidos por la estación meteorológica, $\bar{R}_{si}$ es el promedio de los datos obtenidos por modelo y n es el número total de observaciones [34].

Asimismo, la literatura clasifica el desempeño de los modelos en función de los valores porcentuales de RMSE y MAE de la siguiente manera: excelente ($RMSE < 10\%$), bueno ($10\% \leq RMSE < 20\%$), regular ($20\% \leq RMSE < 30\%$) y deficiente ($RMSE \geq 30\%$). Un criterio similar se aplica al MAE [35], [36].

III. RESULTADOS

La serie temporal de radiación solar en la figura 2 presenta una marcada estacionalidad, caracterizada por ciclos anuales bien definidos, con máximos durante el verano y mínimos en invierno, lo que refleja la influencia directa de la posición solar y la duración del día. Sobre este comportamiento periódico se superpone una variabilidad diaria considerable, atribuida a factores meteorológicos como la nubosidad, aerosoles y condiciones climáticas locales, que generan fluctuaciones de corto plazo. La amplitud entre picos y valles anuales es significativa, evidenciando la fuerte dependencia estacional de la radiación; sin embargo, no se observa una tendencia sostenida a largo plazo.

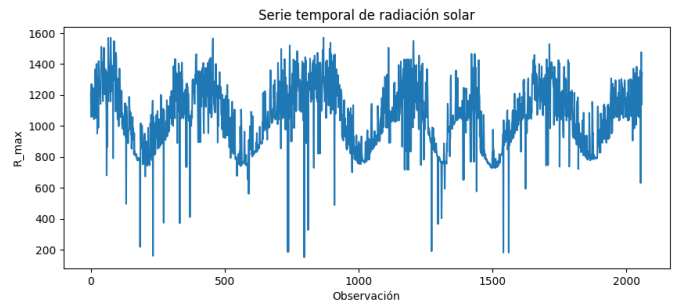


Fig. 2 Valores máximos diarios de la radiación solar.

La figura 3 muestra la serie temporal de la amplitud térmica diaria, definida como la diferencia entre la temperatura máxima y mínima. Se evidencian fluctuaciones significativas a lo largo del periodo de estudio, con valores entre 1.3 °C y 18.3 °C, lo que sugiere un comportamiento estacional asociado a la dinámica atmosférica regional. Los mayores valores se relacionan con condiciones de cielo despejado y baja humedad, mientras que los menores se presentan en periodos más templados o nublados. Dado su impacto en

procesos biofísicos y energéticos, la amplitud térmica constituye una variable climática relevante para estudios predictivos y de cambio climático. Se recomienda complementar el análisis mediante descomposición estacional y evaluación de tendencias para caracterizar su comportamiento temporal.

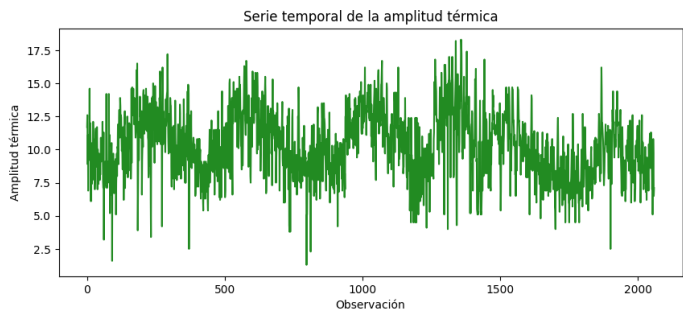


Fig. 3 Valores temporales de amplitud térmica.

A. Distribución de las variables

En la figura 4, se presentan la distribución de las variables: El análisis de las distribuciones muestra que el día juliano presenta una frecuencia uniforme, lo que indica una cobertura temporal equilibrada a lo largo del año. Las temperaturas máxima, mínima y media exhiben comportamientos aproximadamente normales, con valores concentrados alrededor de sus medias y ligeras asimetrías asociadas a periodos más cálidos o fríos. La amplitud térmica presenta una distribución sesgada a la derecha, predominando diferencias térmicas pequeñas o moderadas. Por su parte, la radiación observada muestra una variabilidad amplia con posibles múltiples picos, reflejando la influencia estacional. Finalmente, la radiación extraterrestre evidencia una forma bimodal característica, coherente con los máximos y mínimos anuales asociados a los solsticios, lo que confirma el control astronómico del ciclo de radiación solar.

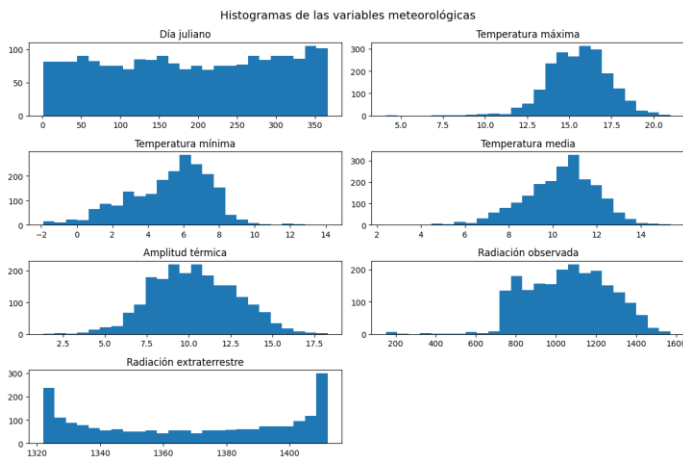


Fig. 4 Distribución de las variables.

B. Matriz de correlación

En la figura 5, el análisis de correlación mostró que la radiación observada (R_{max}) está fuertemente influenciada por la radiación extraterrestre (R_a), presentando una relación positiva elevada, lo que confirma que la disponibilidad de energía solar en el tope de la atmósfera constituye el principal forzamiento físico del régimen radiativo superficial. De manera complementaria, las temperaturas mínima y media evidenciaron correlaciones positivas moderadas, indicando que días con mayor radiación suelen asociarse a condiciones atmosféricas más estables y cielos despejados, favoreciendo un mayor calentamiento del aire tanto diurno como nocturno. En contraste, la amplitud térmica mostró una correlación negativa moderada, lo que sugiere que variaciones térmicas más pequeñas podrían estar vinculadas a la presencia de nubosidad o humedad, factores que limitan la transmisión directa de la radiación solar. La temperatura máxima presentó una relación débil, mientras que el día juliano por sí solo no explicó linealmente la variabilidad de la radiación, ya que su efecto se manifiesta indirectamente a través de la estacionalidad capturada por R_a . Asimismo, las variables térmicas mostraron fuertes interrelaciones entre sí, particularmente entre las temperaturas máxima, mínima y media, evidenciando la coherencia física del conjunto de datos y reflejando patrones climáticos consistentes con la dinámica energética local.

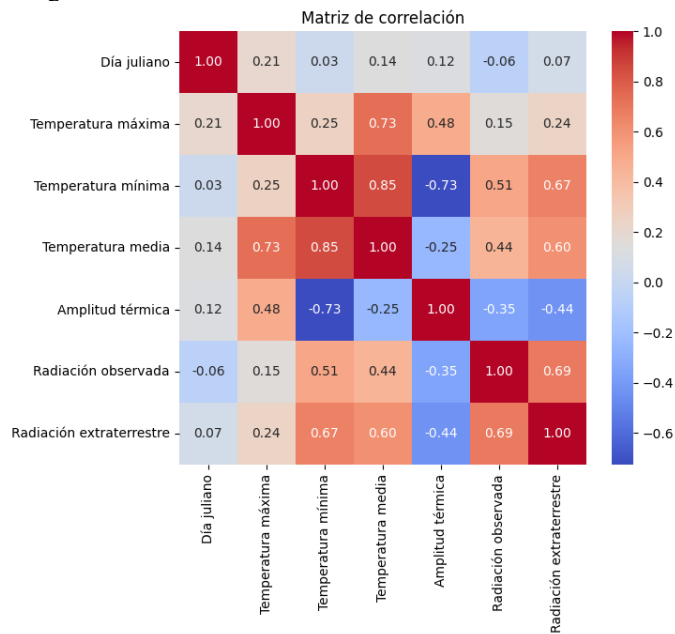


Fig. 5 Matriz de correlación.

C. Resultados de Random Forest

La Figura 6 presenta el desempeño del modelo Random Forest en la predicción de la radiación solar máxima diaria. Se observa que la serie estimada reproduce adecuadamente la tendencia general y la estacionalidad de los valores observados, manteniendo una proximidad consistente entre la curva predicha y la real. No obstante, se identifican

discrepancias más notorias en los extremos locales, particularmente en picos y valles pronunciados, donde el modelo tiende a suavizar las variaciones abruptas, comportamiento característico de los métodos basados en promedios de árboles. Desde el punto de vista cuantitativo, el modelo alcanzó un RMSE de 124.24 W/m² y un MAE de 96.50 W/m², lo que indica errores medios moderados en la estimación diaria. Asimismo, el coeficiente de determinación ($R^2 = 0.5135$) evidencia que el modelo explica aproximadamente el 51.35% de la variabilidad de la radiación observada. En comparación con los demás modelos evaluados, Random Forest presentó el mejor desempeño global, registrando los menores errores y mayor capacidad explicativa, lo que sugiere una adecuada captura de las relaciones no lineales entre las variables meteorológicas y la radiación solar superficial.

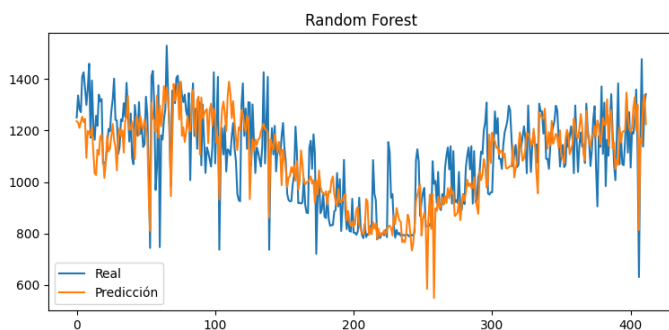


Fig. 6 Random Forest: Real vs predicho.

La Figura 7 El análisis de importancia de variables del modelo Random Forest evidenció que la radiación extraterrestre constituye el predictor dominante de la radiación solar observada, reflejando el control astronómico asociado a la posición relativa de la Tierra respecto al Sol. En segundo lugar, la amplitud térmica diaria mostró una contribución significativa, lo que sugiere que la oscilación térmica es un indicador relevante de las condiciones atmosféricas que favorecen o limitan la transmisión de la radiación solar. El día juliano aportó información adicional, posiblemente vinculada a patrones estacionales no completamente capturados por la radiación extraterrestre. En contraste, las temperaturas máxima, media y mínima presentaron una influencia menor, lo que indica que su efecto sobre la radiación observada es indirecto o se encuentra parcialmente integrado a través de las variables dominantes, evidenciando una jerarquía clara en la relevancia de los predictores dentro del modelo.

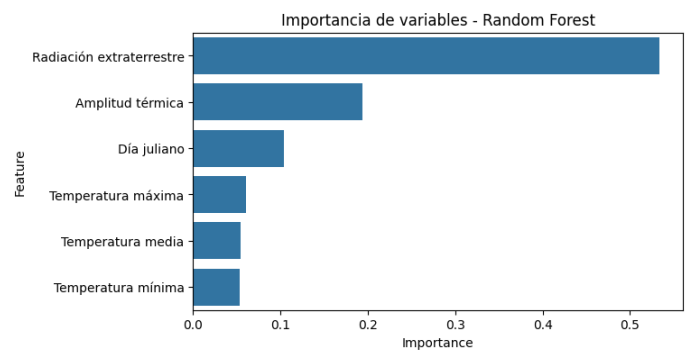


Fig. 7 Random Forest: Importancia de variable.

D. Resultados de XGBoost

La Figura 8 muestra el desempeño del modelo XGBoost en la estimación de la radiación solar máxima diaria. Visualmente, la serie predicha reproduce la tendencia general y el comportamiento estacional de los datos observados; sin embargo, se evidencian mayores dificultades para capturar variaciones abruptas y la magnitud de los valores extremos, particularmente en picos y valles pronunciados. Este comportamiento sugiere una ligera tendencia al suavizado de la señal, lo que incrementa los errores de predicción en periodos de alta variabilidad. En términos cuantitativos, el modelo registró un RMSE de 132.32 W/m² y un MAE de 100.78 W/m², valores superiores a los obtenidos con Random Forest, lo que indica menor precisión promedio. Asimismo, el coeficiente de determinación ($R^2 = 0.4482$) muestra que el modelo explica aproximadamente el 44.82% de la variabilidad de la radiación observada. En consecuencia, aunque XGBoost demostró una capacidad predictiva aceptable y consistente con su naturaleza de modelo de ensamble basado en gradiente, su desempeño fue inferior al de Random Forest en este conjunto de datos, situándose en segundo lugar en términos de exactitud predictiva.

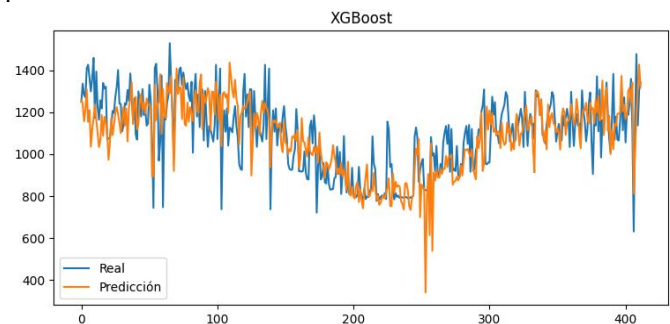


Fig. 8 XGBoost: real vs predicho.

En la figura 9, El análisis de importancia de variables del modelo XGBoost confirmó que la radiación extraterrestre es el predictor predominante, concentrando la mayor contribución explicativa y ratificando su papel como principal forzamiento físico de la radiación superficial. La amplitud térmica diaria se posicionó nuevamente como la segunda variable más relevante, lo que evidencia la influencia consistente de las

condiciones atmosféricas asociadas a la variabilidad térmica en la transmisión de la radiación solar. El día juliano presentó una importancia menor, lo que sugiere que el modelo logra capturar de forma más eficiente la estacionalidad directamente a través de la radiación extraterrestre, reduciendo la redundancia de esta variable temporal. Finalmente, las temperaturas media, mínima y máxima mostraron contribuciones relativamente bajas, indicando que su efecto es indirecto o ya se encuentra parcialmente representado por variables de mayor contenido físico, manteniendo un patrón de influencia similar al observado en Random Forest.

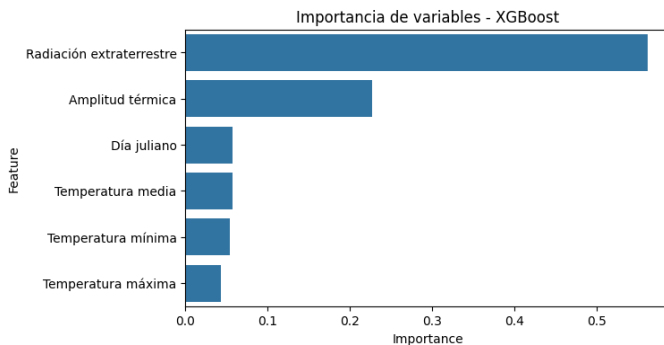


Fig. 9 XGBoost: Importancia de variable.

E. Resultado de LightGBM

La Figura 10 presenta el desempeño del modelo LightGBM en la predicción de la radiación solar máxima diaria. A nivel visual, la serie estimada evidencia una menor concordancia con los valores observados, mostrando discrepancias más pronunciadas tanto en la tendencia general como en la representación de la variabilidad diaria. En particular, el modelo presenta dificultades para reproducir adecuadamente las fluctuaciones de alta frecuencia y los valores extremos, lo que se traduce en una mayor separación entre la curva predicha y la real. Este comportamiento sugiere una capacidad limitada para capturar la complejidad no lineal del fenómeno bajo las configuraciones empleadas. Desde el punto de vista cuantitativo, LightGBM registró un RMSE de 141.63 W/m² y un MAE de 109.35 W/m², correspondientes a los mayores errores entre los modelos evaluados. Asimismo, el coeficiente de determinación ($R^2 = 0.3679$) indica que el modelo explica únicamente el 36.79% de la variabilidad de la radiación observada. En consecuencia, su desempeño predictivo fue inferior al de Random Forest y XGBoost, posicionándose como el modelo de menor capacidad explicativa en este estudio.

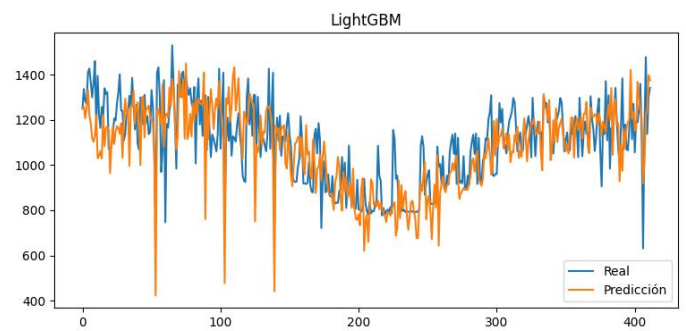


Fig. 10 LightGBM: Real vs predico.

En la figura 11, Aunque no se dispusieron de valores numéricos exactos de importancia para el modelo LightGBM, la inspección visual del gráfico de relevancia de variables evidenció un comportamiento consistente con los demás modelos basados en árboles de decisión. La radiación extraterrestre se identifica como el predictor dominante, lo cual es coherente con su relación física directa con la energía solar incidente, mientras que la amplitud térmica diaria se posiciona como el segundo factor más influyente, reflejando el impacto de las condiciones atmosféricas sobre la variabilidad radiativa. En contraste, el día juliano y las temperaturas máxima, mínima y media presentan contribuciones menores, indicando que su efecto es complementario o indirecto, y que su influencia resulta menos significativa en comparación con las variables de mayor contenido físico.

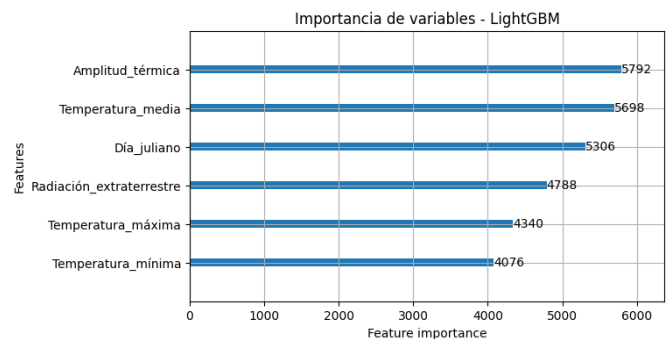


Fig. 11 LightGBM: Importancia de variable.

IV. CONCLUSIONES

Los resultados confirman que la radiación extraterrestre es el principal factor determinante de la variabilidad de la radiación solar superficial en la ciudad de Puno, seguida por la amplitud térmica diaria, lo que evidencia la influencia conjunta de los forzamientos astronómicos y las condiciones atmosféricas locales. Los modelos basados en árboles de decisión demostraron una capacidad adecuada para capturar relaciones no lineales entre las variables meteorológicas y la radiación observada. Entre ellos, Random Forest alcanzó el mejor desempeño predictivo, presentando los menores errores y mayor capacidad explicativa, mientras que XGBoost mostró

un rendimiento intermedio y LightGBM la menor precisión bajo las configuraciones evaluadas. Estos hallazgos indican que los métodos de ensamble constituyen una alternativa robusta para la estimación de radiación solar en zonas con limitada instrumentación radiométrica. En consecuencia, la implementación de modelos de inteligencia artificial puede contribuir significativamente al aprovechamiento de la energía solar, la planificación climática y la gestión ambiental en regiones altoandinas.

REFERENCIAS

- [1] B. Jamil, K. Irshad, A. Algahtani, S. Islam, M. A. Ali, y A. Shahab, «On the calibration and applicability of global solar radiation models based on temperature extremes in India», *Environ. Prog. Sustain. Energy*, vol. 39, n.º 1, pp. 1-13, 2020.
- [2] M. M. Khan, M. J. Ahmad, y B. Jamil, «Development of models for the estimation of global solar radiation over selected stations in India», *Green Energy Technol.*, vol. PartF2, pp. 149-160, 2016.
- [3] M. Huilca, C. W. Taipe, y M. Saavedra, «Índice Ultravioleta en la ciudad de Puno para cielos claros», *Rev. Investig. Altoandinas - J. High Andean Res.*, vol. 19, n.º 2, pp. 211-218, jun. 2017.
- [4] M. Moraes, «Interaction between vitamin D supplementation and sunlight exposure in women living in opposite latitudes (the D-SOL study)», University of Surrey, 2019.
- [5] A. A. Adeola, Z. Huan, y C. C. Enweremadu, «Evaluation of global solar radiation using multiple weather parameters as predictors for South Africa provinces», *Therm. Sci.*, vol. 19, pp. S495-S509, 2015.
- [6] O. T. Kolebaje, A. Ikusika, y P. Akinyemi, «Estimating solar radiation in Ikeja and Port Harcourt via correlation with relative humidity and temperature», *Int. J. Energy Prod. Manag.*, vol. 1, n.º 3, pp. 253-262, 2016.
- [7] A. Muhammad y T. . . Darma, «Estimation of Global Solar Radiation for Kano State Nigeria Based on Meteorological Data», *IOSR J. Appl. Phys.*, vol. 6, n.º 6, pp. 19-23, 2014.
- [8] M. S. Okundamiya, J. O. Emagbetere, y E. A. Ogunjor, «Evaluation of various global solar radiation models for Nigeria», *Int. J. Green Energy*, vol. 13, n.º 5, pp. 505-512, 2016.
- [9] S. Olayinka, «Estimation of global and diffuse solar radiations for selected cities in Nigeria», *Int. J. Energy Environ. Eng.*, vol. 2, n.º 3, pp. 13-33, 2011.
- [10] K. Gairaa y Y. Bakelli, «A Comparative Study of Some Regression Models to Estimate the Global Solar Radiation on a Horizontal Surface from Sunshine Duration and Meteorological Parameters for Ghardaia Site, Algeria», *ISRN Renew. Energy*, vol. 2013, pp. 1-11, 2013.
- [11] Mulandzi ST, Sankarau V, y Lysko MD, «Solar radiation analysis and regression coefficients for the Vhembe Region, Limpopo Province, South Africa», *J. Energy South. Africa*, vol. 24, n.º 3, pp. 02-07, 2013.
- [12] K. Djaman, L. Diop, K. Koudahe, A. Bodian, y P. M. Ndiaye, «Evaluation of temperature-based solar radiation models and their impact on Penman-Monteith reference evapotranspiration in a semiarid climate», *Int. J. Hydrol.*, vol. 4, n.º 2, pp. 84-95, 2020.
- [13] J. L. Bosch, G. López-Rodríguez, F. J. Battles, y J. Tovar-Pescador, «Modelado de la Radiación Solar Global Horaria Sobre Superficie Inclinada Mediante Redes Neuronales Artificiales», *Asoc. Meteorológica Española*, n.º 29, 2006.
- [14] A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan, y Z. Haiou, «Modeling of solar energy systems using artificial neural network: A comprehensive review», *Sol. Energy*, vol. 180, n.º October 2018, pp. 622-639, 2019.
- [15] F. Rodríguez, A. Fleetwood, A. Galarza, y L. Fontán, «Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control», *Renew. Energy*, vol. 126, pp. 855-864, 2018.
- [16] B. Jamil y N. Akhtar, «Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: A case study», *Renew. Sustain. Energy Rev.*, vol. 78, n.º April, pp. 329-355, 2017.
- [17] F. Eraso-Checa, J. B. Jimenez, D. Escobar, y S. Insuasty, «Global radiation estimation using a polynomial function on UV index», *Proc. 2018 IEEE PES Transm. Distrib. Conf. Exhib. - Lat. Am. T D-LA 2018*, 2018.
- [18] P. Paredes y L. S. Pereira, «Computing FAO56 reference grass evapotranspiration PM-ET o from temperature with focus on solar radiation», *Agric. Water Manag.*, vol. 215, n.º December 2018, pp. 86-102, 2019.
- [19] L. Cornejo-Bueno, C. Casanova-Mateo, J. Sanz-Justo, y S. Salcedo-Sanz, «Machine learning regressors for solar radiation estimation from satellite data», *Sol. Energy*, vol. 183, n.º December 2018, pp. 768-775, 2019.
- [20] L. Mazon Aguiar, J. Polo, J. M. Vindel, y A. Oliver, «Analysis of satellite derived solar irradiance in islands with site adaptation techniques for improving the uncertainty», *Renew. Energy*, vol. 135, pp. 98-107, 2019.
- [21] S. Ghimire *et al.*, «Hybrid Convolutional Neural Network-Multilayer Perceptron Model for Solar Radiation Prediction», *Cognit. Comput.*, vol. 15, n.º 2, pp. 645-671, 2023.
- [22] J. N. Latosińska, M. Latosińska, y J. Bielak, «Towards modelling ultraviolet index in global scale. Artificial neural networks approach», *Aerosp. Sci. Technol.*, vol. 41, pp. 189-198, 2015.
- [23] S. S. Prasad, R. C. Deo, S. Salcedo-Sanz, N. J. Downs, D. Casillas-Pérez, y A. V. Parisi, «Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction», *Comput. Methods Programs Biomed.*, vol. 241, 2023.
- [24] S. S. Prasad, R. C. Deo, N. Downs, D. Igoe, A. V. Parisi, y J. Soar, «Cloud Affected Solar UV Prediction with Three-Phase Wavelet Hybrid Convolutional Long Short-Term Memory Network Multi-Step Forecast System», *IEEE Access*, vol. 10, pp. 24704-24720, 2022.
- [25] S. Ghimire, B. Bhandari, D. Casillas-Pérez, R. C. Deo, y S. Salcedo-Sanz, «Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia», *Eng. Appl. Artif. Intell.*, vol. 112, 2022.
- [26] S. Ghimire, R. C. Deo, D. Casillas-Pérez, y S. Salcedo-Sanz, «Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms», *Appl. Energy*, vol. 316, 2022.
- [27] M. S. Naveed *et al.*, «Leveraging advanced AI algorithms with transformer-infused recurrent neural networks to optimize solar irradiance forecasting», *Front. Energy Res.*, vol. 12, n.º October, pp. 1-18, 2024.
- [28] S. Ghimire, B. Bhandari, D. Casillas-Pérez, R. C. Deo, y S. Salcedo-Sanz, «Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia», *Eng. Appl. Artif. Intell.*, vol. 112, n.º April, p. 104860, 2022.
- [29] D. Guijo-Rubio *et al.*, «Evolutionary artificial neural networks for accurate solar radiation prediction», *Energy*, vol. 210, 2020.
- [30] T. Peng, C. Zhang, J. Zhou, y M. S. Nazir, «An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting», *Energy*, vol. 221, 2021.
- [31] M. Guermoui, F. Melgani, K. Gairaa, y M. L. Mekhalfi, «A comprehensive review of hybrid models for solar radiation forecasting», *Journal of Cleaner Production*, vol. 258. Elsevier Ltd, 2020.
- [32] C. W. Taipe, U. Yancachajlla, J. R. Gallegos, y E. G. Mendoza, «Índice de radiación ultravioleta estimado por la Estación Meteorológica Davis de la Universidad Nacional de Juliaca», *Nawparisun*, vol. 2, pp. 37-42, 2020.
- [33] M. Blal *et al.*, «A prediction models for estimating global solar radiation and evaluation meteorological effect on solar radiation potential under several weather conditions at the surface of Adrar environment», *Meas. J. Int. Meas. Confed.*, vol. 152, p. 107348, 2020.
- [34] H. Tabari, «Evaluation of reference crop evapotranspiration equations in various climates», *Water Resour. Manag.*, vol. 24, n.º 10, pp. 2311-2337, 2010.
- [35] M. F. Li, L. Fan, H. Bin Liu, P. T. Guo, y W. Wu, «A general model for estimation of daily global solar radiation using air temperatures and site geographic parameters in Southwest China», *J. Atmos. Solar-Terrestrial Phys.*, vol. 92, pp. 145-150, 2013.
- [36] Yanfeng, «Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China», *Energy Convers. Manag.*, vol. 154, n.º September, pp. 168-179, 2017.