

Solar Radiation Prediction on Photovoltaic Systems Using Machine Learning Techniques

Luis-Eduardo Ordoñez-Palacios; Daniel-Andrés León-Vargas; Víctor-
Andrés Bucheli-Guerrero; Hugo-Armando Ordoñez-Eraso

Citación: L.-E. Ordoñez-Palacios, D.-A. León-Vargas, V.-A. Bucheli-
Guerrero, H.-A. Ordoñez-Eraso, “Solar Radiation Prediction on
Photovoltaic Systems Using Machine Learning Techniques,” *Revista
Facultad de Ingeniería*, vol. 29 (54), e11751, 2020.

<https://doi.org/10.19053/01211129.v29.n54.2020.11751>

Recibido: Julio 12, 2020; **Aceptado:** Septiembre 17, 2020;

Publicado: Septiembre 18, 2020

Derechos de reproducción: Este es un artículo en acceso abierto
distribuido bajo la licencia [CC BY](#)



Conflicto de intereses: Los autores declaran no tener conflicto de
intereses.

Solar Radiation Prediction on Photovoltaic Systems Using Machine Learning Techniques

Luis-Eduardo Ordoñez-Palacios¹

Daniel-Andrés León-Vargas²

Víctor-Andrés Bucheli-Guerrero³

Hugo-Armando Ordoñez-Eraso⁴

Abstract

Estimation of solar radiation is essential to help decision-makers in the planning of isolated solar energy farms or connected to electricity distribution networks to take advantage of renewable energy sources, reduce the impact produced by climate change, and increase coverage rates in electricity service. The number of existing measurement stations is insufficient to cover the entire geography of a region, and many of them are not capturing solar radiation data. Therefore, it is important to use mathematical, statistical, and artificial intelligence models, which allow predicting solar radiation from meteorological data available. In this work, datasets taken from measurement stations located in the cities of Cali and Villavicencio were used, in addition to a dataset generated by the World Weather Online API for the town of Mocoa, to carry out solar radiation estimations using different machine learning techniques for regression and classification to evaluate their performance. Although in most related works researchers used deep learning to predict solar radiation, this work showed that, while artificial neural networks are the most widely used

¹ Universidad del Valle (Santiago de Cali-Valle del Cauca, Colombia). luis.ordonez.palacios@correounivalle.edu.co. ORCID: [0000-0001-5154-9472](https://orcid.org/0000-0001-5154-9472)

² M. Sc. Universidad del Valle (Santiago de Cali-Valle del Cauca, Colombia). daniel.leon@correounivalle.edu.co. ORCID: [0000-0002-8283-2243](https://orcid.org/0000-0002-8283-2243)

³ Ph. D. Universidad del Valle (Santiago de Cali-Valle del Cauca, Colombia). victor.bucheli@correounivalle.edu.co. ORCID: [0000-0002-0885-8699](https://orcid.org/0000-0002-0885-8699)

⁴ Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). hugoordonez@unicauca.edu.co. ORCID: [0000-0002-3465-5617](https://orcid.org/0000-0002-3465-5617)

technique, other machine learning algorithms such as Random Forest, Vector Support Machines and AdaBoost, also provide estimates with sufficient precision to be used in this field of study.

Keywords: deep learning; machine learning; photovoltaic systems; prediction model; solar radiation; supervised learning.

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

Resumen

La estimación de la radiación solar es fundamental para quienes participan en la planificación de granjas de energía solar, ya sean aisladas o conectadas a las redes de distribución eléctrica. Esto para el aprovechamiento de las fuentes de energía renovables, reducir el impacto producido por el cambio climático, e incrementar los índices de cobertura en el servicio eléctrico. De igual manera, el número de estaciones de medición existentes es insuficiente para cubrir toda la geografía de una región, y muchas de ellas no están capturando datos de radiación solar. Por consiguiente, es importante hacer uso de modelos matemáticos, estadísticos y de inteligencia artificial que permitan predecir la radiación solar a partir de datos meteorológicos disponibles. En este trabajo se utilizaron conjuntos de datos tomados de estaciones de medición ubicadas en las ciudades de Cali y Villavicencio, además de un conjunto de datos generado por la API World Weather Online para la ciudad de Mocoa. La razón fue realizar estimaciones de radiación solar utilizando distintas técnicas de aprendizaje automático para regresión y clasificación; el principal objetivo fue evaluar su desempeño. Aunque en la mayoría de los trabajos relacionados los investigadores utilizaron el aprendizaje profundo para la predicción de la radiación solar, este estudio demostró que, si bien las redes neuronales artificiales son la técnica más utilizada, otros algoritmos de aprendizaje automático como Random Forest, Máquinas de Soporte Vectorial y AdaBoost también proporcionan estimaciones con suficiente precisión para ser utilizados en este campo de estudio.

Palabras clave: aprendizaje automático; aprendizaje profundo; aprendizaje supervisado; modelo de predicción; radiación solar; sistemas fotovoltaicos.

Predição de radiação solar em sistemas fotovoltaicos utilizando técnicas de aprendizagem automática

Resumo

A estimativa da radiação solar é fundamental para quem participa na planificação de granjas de energia solar, já sejam isoladas ou conectadas às redes de distribuição elétrica. Isto para o aproveitamento das fontes de energia renováveis, reduzir o impacto produzido pela mudança climática, e incrementar os índices de cobertura no serviço elétrico. De igual maneira, o número de estações de medição existentes é insuficiente para cobrir toda a geografia de uma região, e muitas delas não estão capturando dados de radiação solar. Por conseguinte, é importante fazer uso de modelos matemáticos, estatísticos e de inteligência artificial que permitam prever a radiação solar a partir de dados meteorológicos disponíveis. Neste trabalho utilizaram-se conjuntos de dados tomados de estações de medição localizadas nas cidades de Cali e Villavicencio, além de um conjunto de dados gerado pela API World Weather Online para a cidade de Mocoa. A razão foi realizar estimativas de radiação solar utilizando distintas técnicas de aprendizagem automática para regressão e classificação; o principal objetivo foi avaliar seu desempenho. Ainda que na maioria dos trabalhos relacionados os pesquisadores utilizaram a aprendizagem profunda para a predição da radiação solar, este estudo demonstrou que, se bem as redes neuronais artificiais são a técnica mais utilizada, outros algoritmos de aprendizagem automática como Random Forest, Máquinas de Suporte Vectorial e AdaBoost também proporcionam estimativas com suficiente precisão para serem utilizados neste campo de estudo.

Palavras chave: aprendizagem automática; aprendizagem profunda; aprendizagem supervisada; modelo de predição; radiação solar; sistemas fotovoltaicos.

I. INTRODUCCIÓN

Con el agotamiento escalonado en las reservas de combustibles fósiles [1], el calentamiento global y los cambios climáticos extremos [2], la mayoría de los países está tomando conciencia acerca de la importancia de utilizar tecnologías innovadoras para la generación de energía limpia, renovable y amigable con el ambiente. Entre estas tecnologías, los sistemas fotovoltaicos se constituyen como una alternativa que se ha fortalecido recientemente, en parte, gracias a que los costos se han reducido gradualmente en los últimos años y el desarrollo tecnológico ha mejorado la eficiencia de sus componentes [3]. Por otro lado, aun cuando la energía solar se caracteriza por ser inagotable y limpia, su explotación y transformación a energía eléctrica depende de factores climáticos, geográficos y técnicos.

La instalación de sistemas fotovoltaicos requiere de una evaluación conceptual previa con el objetivo de conocer la viabilidad del sitio tentativo respecto a las necesidades que se quieren satisfacer. Esta evaluación previa es clave para apoyar a las personas que toman las decisiones en la planificación e implementación de granjas de energía solar (que bien pueden ser aisladas o conectadas a las redes de distribución eléctrica), así como en la operación y la administración de la energía producida. Todo esto se suma al creciente interés mundial por el desarrollo de sistemas de generación de energía limpios y amigables con el ambiente.

Es esencial conocer el comportamiento de la radiación solar global de un lugar específico y, al mismo tiempo, es trascendental en el cálculo del número de módulos solares y los demás componentes que se requieren para producir cierta cantidad de energía eléctrica. Sin embargo, los datos de irradiancia no están disponibles en la mayoría de las estaciones de medición existentes, lo que conlleva a precisar de tiempo adicional para realizar las mediciones y el procesamiento de los datos, ocasionando retrasos y riesgos en la ejecución de este tipo de proyectos.

Actualmente, la Inteligencia Artificial [4] está ampliando las fronteras de la predicción en campos como las finanzas, la justicia [5], la salud, el clima, la agricultura, las ventas, el transporte, entre otros; esto gracias al aprendizaje automático, el procesamiento de imágenes, el procesamiento del lenguaje natural, la robótica, etc.

En el sector de la generación de energía eléctrica, las técnicas del aprendizaje automático han cobrado gran importancia en la estimación de atributos meteorológicos a partir de conjuntos de datos estructurados, semiestructurados y no estructurados. Aunque muchas de las estaciones no están tomando lecturas de brillo solar, existen algunas que sí la miden, junto con otras variables climáticas, con el fin de lograr una mejor comprensión del ambiente. En este trabajo se verifica si estas variables son suficientes para estimar la energía luminosa procedente del sol. El presente estudio logró confirmar que la literatura relacionada con la estimación de la radiación solar en sistemas fotovoltaicos utilizando técnicas de Inteligencia Artificial se produjo en los últimos diez años, en parte, gracias al gran impulso que han tenido las Ciencias de la Computación [6]. También se pudo determinar que la mayoría de los investigadores modernos utilizan el aprendizaje profundo para realizar estimaciones de radiación solar, como se ha presentado en trabajos preliminares [7-12]. A pesar de ello, otros documentos han usado técnicas como árboles de decisión, regresión lineal, no lineal, multivariante y logística [13]; de igual manera, se han utilizado algoritmos como Random Forest [14], clustering [15], máquinas de soporte vectorial [16-17] y modelos híbridos [18].

Sharadga et al. [19] emplean modelos de predicción basados en métodos estadísticos y de inteligencia artificial para la estimación de salida de energía eléctrica a partir de la radiación solar en sistemas fotovoltaicos. Con respecto a la inteligencia artificial, se consideraron los siguientes modelos de Redes Neuronales: Bi-LSTM, agrupamiento Fuzzy c-Means, LSTM, LRNN, MLP y feedforward NNs. De acuerdo con los resultados, el modelo Bi-LSTM tuvo el mejor desempeño, con un coeficiente de correlación del 98%, mientras que RMSE tuvo uno de 0.791, aunque el modelo perceptrón multicapa (MLP) fue el algoritmo más rápido. El estudio demostró que las redes neuronales son más precisas que los modelos estadísticos y requieren menos tiempo de cálculo.

Las redes neuronales y el aprendizaje automático son bastante utilizados en aplicaciones de radiación solar, evaluando el desempeño del perceptrón multicapa (MLP) y los árboles de decisión potenciados mediante la combinación con regresión lineal para la estimación de la energía solar en el sur de Argelia. El conjunto de

datos utilizado incluye las variables radiación solar global diaria, radiación solar global extraterrestre, temperatura del aire y duración de la luz solar; los resultados demuestran que el modelo MLP tuvo un mejor rendimiento según los indicadores coeficiente de determinación R^2 de 97.7% y el RMSE de 0.033 [20].

Otros estudios realizan una mejora en la técnica de regresión Random Forest para predecir la radiación solar en entornos complejos, considerando los factores que influyen en la energía solar, y realizando la selección de parámetros óptimos a partir del análisis de error OOB. Se demostró la validez del modelo gracias a la reducción del error cuadrático medio hasta 19.6085 y la mejora en la precisión de la estimación con un coeficiente de determinación R^2 de 0.9665. Se utilizaron datos de monitoreo de la radiación solar del laboratorio del Instituto de Ingeniería de Shenyang, para una muestra entre 2004 y 2014, que incluye horas de sol, presión atmosférica promedio, velocidad promedio del viento, temperatura promedio del aire, ángulo de elevación solar y humedad promedio [14].

Koo et al. [21] abordan los desafíos relacionados con los patrones espaciales complejos en el promedio mensual de radiación solar diaria en China, así como la carencia de estaciones de observación de esta variable debido a los altos costos de los equipos. El estudio desarrolló un enfoque de estimación mediante técnicas de aprendizaje automático como el algoritmo de agrupamiento K-means y un modelo avanzado de razonamiento basado en casos CBR. Se obtuvo una precisión de estimación promedio de 93.23%, a partir de información recopilada de 97 ciudades de China, en un periodo de 10 años entre 2006 y 2015.

De acuerdo con estudios realizados por Hussain y Alilili [22], los datos solares no están disponibles en todos los puntos de interés a causa de la insuficiencia de estaciones de medición y sensores solares, por ello, la irradiancia debe estimarse utilizando modelos. Los investigadores desarrollaron una técnica híbrida para mejorar el rendimiento de una red neuronal (ANN), analizando cuatro arquitecturas distintas y utilizando el algoritmo de transformación de wavelet discreta. En el proceso de aprendizaje, los modelos emplearon temperatura, humedad relativa, velocidad del viento y duración del sol para estimar la irradiancia horizontal global sobre Abu Dhabi, Emiratos Árabes Unidos. Los resultados mostraron una mejora

importante en el rendimiento de las ANN con un R^2 de 6.84% para el perceptrón multicapa, y un RMSE de 2.78% para redes neuronales de regresión generalizada.

II. METODOLOGÍA

En esta sección se definen las preguntas de interés que condujeron a la construcción del modelo para la estimación de la Radiación Solar utilizando técnicas de Aprendizaje Automático. También se describen las fuentes de información y la manera en cómo se procesaron los datos.

A. Preguntas de interés

Este trabajo se produce a partir de la necesidad existente en el aprovechamiento de fuentes de energía limpia, como el caso de la energía solar; por lo tanto, es trascendental conocer el comportamiento de los niveles de luminosidad solar para la implementación de sistemas fotovoltaicos en lugares específicos de la geografía colombiana. Sin embargo, las mediciones de brillo solar aún son inexistentes en muchas regiones; esto implica la construcción de modelos de Inteligencia Artificial basados en Aprendizaje Automático para predecir la radiación solar a partir de variables climáticas, pero, ¿esas variables son suficientes para realizar una buena estimación de radiación solar? ¿Cuáles son las técnicas con mejor desempeño en la predicción? Y, ¿cuál es el grado de confianza en dichas estimaciones? Estas interrogantes serán despejadas durante el desarrollo del documento.

B. Fuentes de información

En este estudio se emplearon conjuntos de datos públicos de las estaciones de medición de calidad del aire (Escuela República de Argentina, La Flora, Pance, Compartir) del Departamento Administrativo de Gestión del Medio Ambiente (DAGMA) de la Alcaldía de Cali. También se utilizó un conjunto de datos de la estación Catumare, ubicada en la comuna ocho de la ciudad de Villavicencio, como parte del centro de control de Cormacarena, y un conjunto de datos generado por la API World Weather Online para la ciudad de Mocoa. Todos los conjuntos de datos públicos incluyen observaciones históricas de variables como velocidad y dirección

del viento, temperatura, humedad, lluvia y radiación solar, tomadas por hora durante un periodo de tiempo. En contraste, el conjunto de datos de Mocoa no contiene radiación solar. La Tabla 1 presenta los metadatos de cada conjunto de datos utilizado.

Tabla 1. Descripción de los conjuntos de datos.

Id	Estación	Periodo	Variabes	Registros	Uso en el trabajo
1	Escuela República de Argentina	28/01/2018 01/01/2019	12	78241	Entrenamiento de los algoritmos
2	La Flora	06/05/2010 01/01/2019	12	75869	Evaluación del modelo
3	Pance	27/02/2013 01/01/2019	8	51216	
4	Compartir	30/01/2014 01/01/2019	9	43118	
5	Catumare	25/06/2018 31/12/2018	8	4455	
6	Mocoa	01/01/2010 31/12/2019	25	87648	

Se definió la radiación solar como variable objetivo; el mes, la hora, la velocidad y dirección del viento, la temperatura, la humedad y la lluvia se definieron como variables independientes a partir de las cuales se producirán las estimaciones de irradiancia. Considerando que en todos los conjuntos de datos utilizados se conocen los valores de radiación solar, excepto en el conjunto de datos de Mocoa, es preciso subrayar que se seleccionaron técnicas de aprendizaje supervisado como uno de los tipos de Aprendizaje Automático para regresión y clasificación.

C. Algoritmos de aprendizaje automático

En la regresión se emplearon árboles de regresión, regresión lineal múltiple, regresión de soporte vectorial, redes neuronales, Random Forest y Gradient Boosting; para la clasificación se utilizaron árboles de decisión, máquinas de soporte vectorial y AdaBoost. En cuanto a las redes neuronales, se utilizó el perceptrón multicapa para regresión y se probaron diferentes configuraciones, determinando como óptimos los siguientes hiperparámetros: 4 capas ocultas de 100, 75, 50 y 25 neuronas; función de activación ReLU; algoritmo de optimización Adam; término de regularización L2 de 0.00001; tasa de aprendizaje adaptativa, y un número máximo

de 10000 iteraciones. Se utilizó regresión lineal simple para evaluar la correlación entre las variables independientes y la variable objetivo, determinando que la temperatura tiene más reciprocidad con respecto a la radiación solar.

D. Filtrado de datos

El conjunto de datos de la estación Escuela República de Argentina fue utilizado para el entrenamiento del modelo debido a que se encuentra ubicada en el centro de la ciudad de Cali; por ende, se convierte en punto de referencia con respecto a las demás estaciones, las cuales presentan variaciones en las mediciones debido a su ubicación (distancia que las separa). Además, esta estación cuenta con la mayor cantidad de registros. Según datos obtenidos de Google Maps, la estación Escuela República de Argentina está a 3.74 km en línea recta de la estación La Flora, en el norte de la ciudad; a 18.45 km de la estación Pance, en el sur de la ciudad en la zona rural, y a 6.98 km de la estación Compartir, ubicada en el oriente de la ciudad. La Figura 1 muestra la ubicación de las estaciones con respecto a la estación de referencia. El conjunto de datos de la estación Catumare y el de la ciudad de Mocoa son importantes para este estudio debido a que cuentan con características climáticas diferentes, lo que permite evaluar la precisión del modelo en la estimación de la luminosidad proyectada por los rayos solares en otras regiones del país.

Es importante mencionar que los conjuntos de datos presentan discontinuidad (datos faltantes o espacios en blanco) debido a las anomalías que se dan en las estaciones de monitoreo tales como: fallas en los equipos, falta de energía eléctrica en la zona, hurto de equipos o cableado, mantenimiento o cambio de equipos, entre otros. También está la inclusión o exclusión de algunos contaminantes o variables meteorológicas según criterio de los expertos y características de la zona a monitorear.

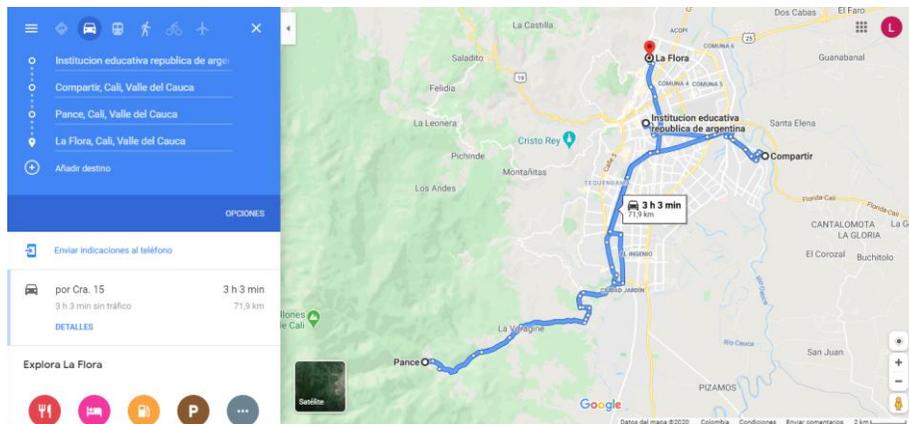


Fig 1. Ubicación de las estaciones de medición. Tomada de Google Maps.

El preprocesamiento de los datos se inicia con la exclusión de las variables de calidad del aire, debido a que no todas las estaciones de monitoreo del clima incluyen estas características, y considerando que los niveles de contaminación del aire en entornos distintos a las ciudades principales son muy bajos o incluso nulos. Además, este estudio pretende estimar la energía luminosa proveniente del sol a partir de datos meteorológicos. Posteriormente, se divide la variable FechaHora en las columnas mes y hora para analizar el comportamiento de la luz solar a lo largo de un día o de un año; por otro lado, se descarta el año y el día de la medición realizada, ya que no se establecen como esenciales para el presente estudio.

Con respecto a la intermitencia en los datos a causa de las razones antes mencionadas, los valores faltantes fueron completados con el promedio entre el valor anterior y el siguiente en cada columna, después se eliminaron aquellos registros que aún contenían valores nulos. Por último, en el caso de la regresión, los datos originales se normalizaron con el método Min-Max, transformando los valores en un rango entre cero y uno. La Tabla 2 muestra la cantidad de registros restantes luego de la eliminación de las observaciones incompletas.

En el modelo de clasificación se transformó la variable radiación solar en las clases 0 y 1; 0 para las observaciones inferiores a 300 Watts/m² y 1 para las demás. El valor definido se tomó como punto de referencia para la transformación de la luz solar en energía eléctrica, considerando la cantidad de paneles y el tamaño de la superficie necesaria para la instalación de sistemas fotovoltaicos. El conjunto de

datos quedó dividido en 26.6% en 1 y 73.4% en 0, dado que, cuando el sol está oculto, la irradiancia es cero, y en las primeras y las últimas horas del día, la radiación solar es baja. Debido a que es recomendable trabajar con datos balanceados, en la clasificación se utilizó la técnica de submuestreo Random Under-sampling para ajustar la distribución de las clases. Así se buscó evitar el *sobreentrenamiento* del algoritmo que conduce al sobreajuste de los datos, quedando finalmente con un conjunto de datos de 9,045 observaciones en cada clase, para un total de 18,090 registros.

Tabla 2. Registros luego de la eliminación.

Id	Estación	Registros antes	Registros después	Pérdida de datos (%)
1	Escuela República de Argentina	78241	33951	56.61
2	La Flora	75869	37411	50.69
3	Pance	51216	304	99.41
4	Compartir	43118	31896	26.03
5	Catumare	4455	4455	0.00
6	Mocoa	87648	87648	0.00

III. RESULTADOS

Posterior al preprocesamiento de los datos, se realizó un re-muestreo en el conjunto de datos de la Escuela República de Argentina, con un 70% para entrenamiento y el 30% restante para evaluar el rendimiento de los algoritmos empleados en el estudio. Luego, se aplicaron distintas técnicas de aprendizaje supervisado para la estimación de la radiación solar. Inmediatamente después del entrenamiento y valoración de los algoritmos, se aplicó la técnica con mayor rendimiento a los datos de prueba de las demás estaciones, para evaluar el desempeño del modelo mediante la comparación entre las predicciones realizadas y las observaciones reales registradas en cada conjunto de datos.

A. Arquitectura

Este apartado presenta el tratamiento aplicado a todos los conjuntos de datos empleados por los diferentes algoritmos de aprendizaje automático, con el propósito de realizar las estimaciones de energía solar mediante regresión y clasificación. La

Figura 2 expone el flujo de la información desde las fuentes hasta la generación de las predicciones en cada una de las etapas del modelo.

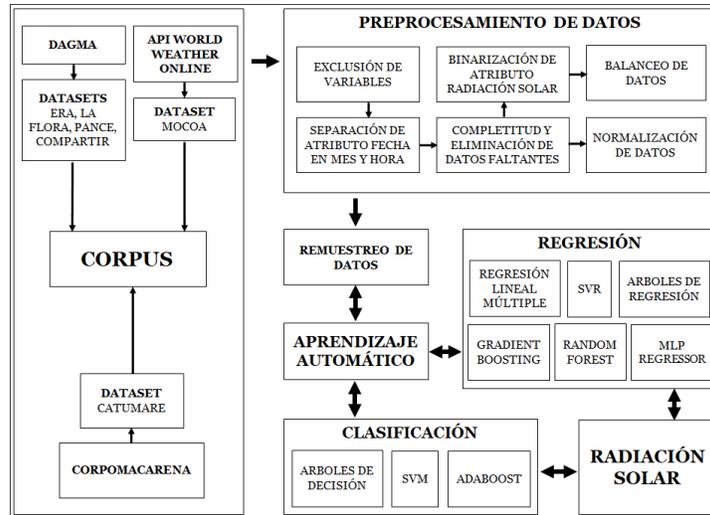


Fig. 2. Flujo de información del modelo.

B. Regresión

En la Tabla 3 se observa que con el algoritmo MLP Regressor se obtuvieron los mejores resultados, según el coeficiente de determinación R^2 en los datos de desarrollo y en la raíz del error cuadrático medio RMSE.

Tabla 3. Resultados obtenidos mediante regresión.

Técnica	R^2 -score entrenamiento	R^2 -score desarrollo	RMSE
Regresión lineal múltiple	0.56	0.55	163.61
Máquinas de Soporte Vectorial para Regresión (SVR)	0.77	0.77	117.73
Árboles de regresión	0.90	0.89	81.66
Gradient Boosting	0.98	0.89	82.26
Random Forest	0.98	0.89	79.99
MLP Regressor	0.90	0.90	77.37

C. Clasificación

La Tabla 4 presenta los resultados obtenidos con los algoritmos de clasificación, resaltando que el método de ensamble AdaBoost tuvo mejor desempeño según las métricas de la matriz de confusión (Accuracy, Precision, Recall y F1-score), tanto para los datos de entrenamiento como para los datos de desarrollo. En contraste,

las máquinas de soporte vectorial superaron al algoritmo AdaBoost en Accuracy en los datos de entrenamiento y en Precisión en los datos de desarrollo.

Tabla 4. Resultados obtenidos mediante clasificación.

Técnica	Entrenamiento	Desarrollo			
	Accuracy	Accuracy	Precisión	Recall	F1-score
Árboles de decisión	0.93	0.93	0.89	0.98	0.93
Máquinas de Soporte Vectorial para Clasificación (SVM)	0.98	0.90	0.91	0.88	0.90
AdaBoost	0.94	0.94	0.90	0.99	0.94

D. Evaluación del modelo

En las Tablas 5 y 6 se pueden apreciar los resultados obtenidos en las predicciones mediante las técnicas con mejor desempeño: Random Forest, MLP Regresor y AdaBoost.

Tabla 5. Resultados obtenidos con Random Forest y MLP Regresor.

Conjunto de datos	Cantidad de datos	Random Forest		MLP Regresor	
		R ² -score	RMSE	R ² -score	RMSE
La Flora	37411	0.88	79.19	0.88	59.84
Pance	304	0.84	51.59	0.82	55.24
Compartir	31896	0.37	114.22	0.44	82.95
Catumare	4455	0.25	154.70	0.74	90.40

Tabla 6. Resultados obtenidos con la técnica de clasificación AdaBoost.

Conjunto de datos	Datos balanceados	Accuracy	Precisión	Recall	F1-score
La Flora	18550	0.94	0.90	0.99	0.94
Pance	66	0.99	0.97	1.00	0.99
Compartir	8902	0.87	0.79	1.00	0.88
Catumare	1532	0.88	0.84	0.95	0.89
Mocoa	63244	0.81	1.00	0.62	0.77

Las Figuras 3 y 4 muestran los diagramas de dispersión obtenidos por la regresión, a partir de la radiación solar real, y estimada en los conjuntos de datos de las estaciones: La Flora, Pance, Compartir y Catumare. Para el conjunto de datos de Mocoa no fue posible evaluar el modelo de regresión, ni tampoco construir la gráfica, dado que no incluye valores de radiación solar.

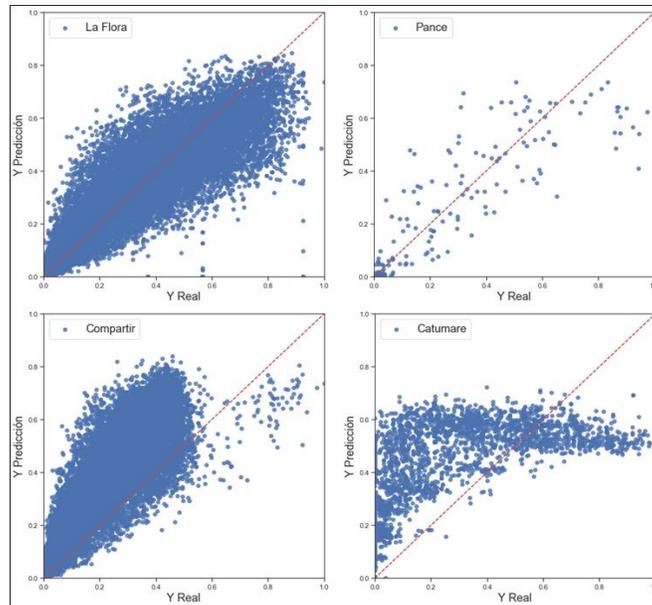


Fig. 3. Radiación solar real y estimada con Random Forest.

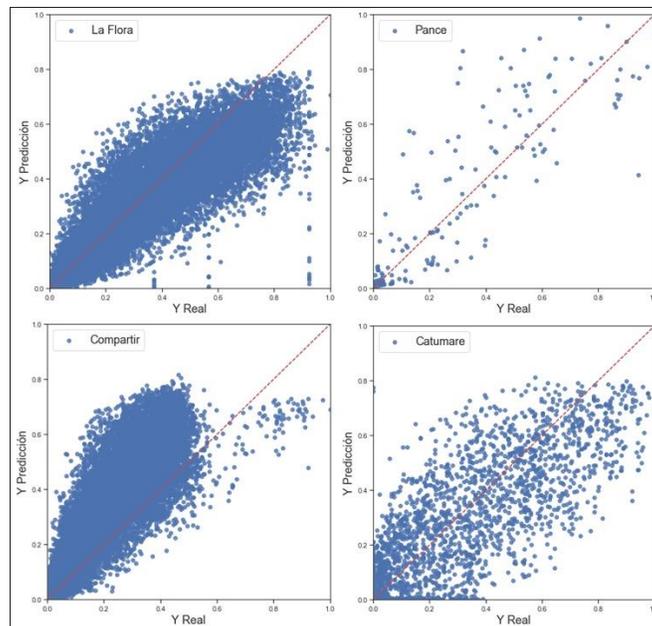


Fig. 4. Radiación solar real y estimada con MLP Regressor.

E. Estimación de radiación solar en la ciudad de Mocoa

La Figura 5 muestra el comportamiento de la radiación solar estimada para la ciudad de Mocoa, la Figura 6 presenta el promedio mensual de radiación solar entre las 6 am y las 4 pm, en el periodo comprendido entre los años 2010 y 2019. Finalmente,

la Figura 7 permite la visualización de las predicciones referentes a la cantidad de horas mensuales con valores de radiación solar mayores o iguales a 300 W/m^2 .

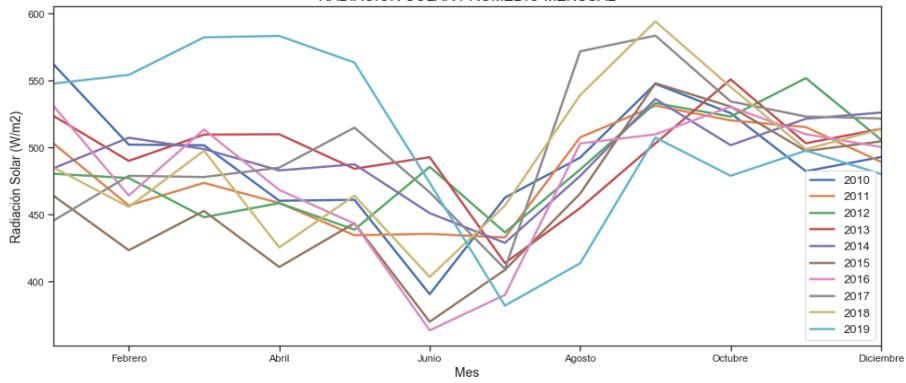


Fig. 5. Radiación solar promedio mensual por año.

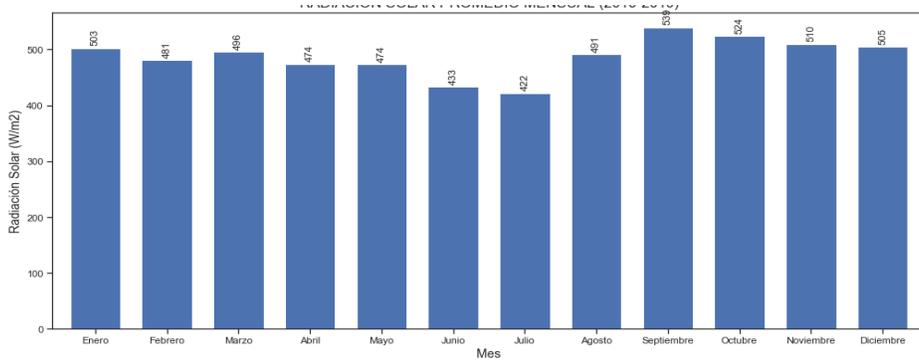


Fig. 6. Radiación solar promedio mensual entre 2010 y 2019.

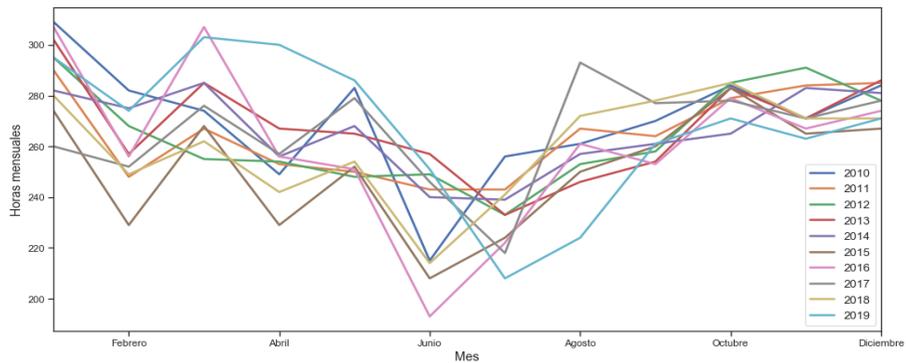


Fig. 7. Horas mensuales con radiación solar mayor o igual a 300 W/m^2 .

IV. DISCUSIÓN

El objetivo de los algoritmos de aprendizaje automático consiste en permitir que las computadoras aprendan. En la actualidad, esta tendencia se ha incrementado en gran medida y es posible observarla en distintos campos como el de la generación de energía; por ende, este estudio ha empleado varias técnicas de aprendizaje supervisado, las cuales han mostrado buen desempeño. Por otro lado, es importante destacar que muchos de los artículos revisados usan redes neuronales artificiales para predecir la energía luminosa que se desprende del sol; por consiguiente, se recomienda continuar este trabajo con la construcción de un modelo robusto, basado en alguno de los distintos marcos de trabajo de aprendizaje profundo, para estimar la irradiancia en cualquier lugar del planeta.

Un modelo de aprendizaje profundo implica la adquisición de muchos más conjuntos de datos de diferentes localizaciones geográficas para incrementar, aún más, la confianza en las estimaciones históricas y el pronóstico del comportamiento futuro en semanas, meses e incluso años. Por otra parte, en este estudio no se tuvieron en cuenta algunos factores que determinan los diferentes tipos de climas regionales como la latitud, la altitud, el relieve, etc., por ello, se considera de suma importancia una investigación más profunda sobre estos aspectos.

Es importante destacar que a todos los conjuntos de datos se les dio el mismo tratamiento en el preprocesamiento de los datos, a pesar de ello, se esperaba un mejor desempeño en la evaluación del modelo para el caso del conjunto de datos de la estación Compartir, dado que su ubicación se encuentra a menos de 7 km de la estación Escuela República de Argentina y ambas se encuentran dentro de la misma ciudad. En el caso de la estación Catumare, también se observa bastante variabilidad en las estimaciones, aunque era algo que se esperaba, debido a que se encuentra situada en otra región a menor altura sobre el nivel del mar.

V. CONCLUSIONES

Comprender el comportamiento de la Radiación Solar en una región determinada es indispensable para la construcción de parques solares correctamente

dimensionados, de modo que conduzcan al máximo aprovechamiento de la energía solar. En este trabajo se logró obtener un buen desempeño en la estimación de la luminosidad del sol a partir las características meteorológicas disponibles en los conjuntos de datos empleados. En el caso del modelo de regresión, la temperatura es la variable que tiene mayor correlación con la radicación solar, mientras que, en el modelo de clasificación, el atributo más importante es la hora del día.

Aunque en la actualidad hay una gran demanda en el uso del aprendizaje profundo para realizar predicciones en diferentes áreas, existe una amplia variedad de técnicas de aprendizaje automático con alto desempeño; por consiguiente, los métodos de ensamble combinan grupos de modelos para incrementar el desempeño en las predicciones. En este trabajo, los mejores resultados para el caso de la regresión se obtuvieron con los algoritmos MLP Regressor, con un coeficiente de determinación R^2 de 0,90 y un RMSE de 77.37, y Random Forest, con un coeficiente de determinación R^2 de 0,89 y un RMSE de 79.99. Por otra parte, en la clasificación, la técnica AdaBoost tuvo mejor desempeño, aun así, las máquinas de soporte vectorial mostraron un rendimiento superior en algunas de las métricas de la matriz de confusión.

Esta investigación propone una evaluación conceptual en la instalación de sistemas fotovoltaicos, a partir de la comprensión de los valores de radiación solar estimados por modelos de inteligencia artificial, como apoyo a las personas encargadas de tomar decisiones. Para el caso de Mocoa, podemos decir que se produce, en promedio, una radiación solar anual de 487.67 W/m^2 y de 6 a 10 horas diarias con radiación solar mayor a 300 W/m^2 , contemplando un margen de error de entre el 10% y el 60% en las estimaciones, basados en los modelos que utilizan conjuntos de datos donde sí existe información de energía solar para confrontarlos.

CONTRIBUCIÓN DE LOS AUTORES

Luis-Eduardo Ordoñez-Palacios: Conceptualización, Análisis Formal, Investigación, Metodología, Escritura - borrador original, Escritura - revisión y edición.

Daniel-Andrés León-Vargas: Análisis Formal, Investigación, Escritura - revisión y edición.

Víctor-Andrés Bucheli-Guerrero: Análisis Formal, Supervisión, Validación, Escritura - revisión y edición.

Hugo-Armando Ordoñez-Eraso: Análisis Formal, Supervisión, Validación, Escritura - revisión y edición.

REFERENCIAS

- [1] A. Doval Adán, "Los grandes problemas geopolíticos del desarrollo mundial: hacia una planificación global del planeta," in *XVIII Congreso de Geógrafos Españoles* 2003. <https://minerva.usc.es/xmlui/handle/10347/20719>
- [2] Z. Li, "Global Warming: Causes and Effects," in *Southern California Conferences for Undergraduate Research*, 2019. https://www.sccur.org/sccur/fall_2019_conference/poster_session_4/131
- [3] M. R. Gámez, A. V. Pérez, A. M. V. Quiroz, and W. M. S. Arauz, "Mejora de la calidad de la energía con sistemas fotovoltaicos en las zonas rurales," *Revista Científica*, vol. 3 (33), pp. 265-274, 2018. <https://doi.org/10.14483/23448350.13104>
- [4] F. G. Cozman, "O futuro da (pesquisa em) inteligência artificial: algumas direções," *Revista USP*, n. 124, pp. 11-20, 2020. <https://doi.org/10.11606/issn.2316-9036.v0i124p11-20>
- [5] H. Ordóñez, C. Cobos, and V. Bucheli, "Modelo de machine learning para la predicción de las tendencias de hurto en Colombia," *Revista Ibérica de Sistemas y Tecnologías de la Información*, vol. 29, pp. 494-506, 2020.
- [6] M. Haenlein, and A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *California Management Review*, vol. 61 (4), pp. 5-14, 2019. <https://doi.org/10.1177/0008125619864925>
- [7] N. Premalatha, and A. Arasu, "Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms", *Journal of Applied Research and Technology*, vol. 14 (3), 2016. <https://doi.org/10.1016/j.jart.2016.05.001>
- [8] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. A. Rahim, and W. A. Khan, "The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review," *Journal of Cleaner Production*, vol. 104, pp. 1-12, 2015. <https://doi.org/10.1016/j.jclepro.2015.04.041>
- [9] A. K. Yadav, and S. S. Chandel, "Solar radiation prediction using Artificial Neural Network techniques: A review," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 772-781, 2014. <https://doi.org/10.1016/j.rser.2013.08.055>
- [10] M. Ozgoren, M. Bilgili, and B. Sahin, "Estimation of global solar radiation using ANN over Turkey," *Expert Systems with Applications*, vol. 39 (5), pp. 5043-5051, 2012. <https://doi.org/10.1016/j.eswa.2011.11.036>
- [11] A. Koca, H. F. Oztop, Y. Varol, and G. O. Koca, "Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey," *Expert Systems with Applications*, vol. 38 (7), pp. 8756-8762, 2011. <https://doi.org/10.1016/j.eswa.2011.01.085>
- [12] A. Rahimikhoob, "Estimating global solar radiation using artificial neural network and air temperature data

- in a semi-arid environment,” *Renewable Energy*, vol. 35 (9), pp. 2131-2135, 2010. <https://doi.org/10.1016/j.renene.2010.01.029>
- [13] D. Li, W. Chen, S. Li, and S. Lou, “Estimation of Hourly Global Solar Radiation Using Multivariate Adaptive Regression Spline (MARS) – A Case Study of Hong Kong,” *Energy*, vol. 186, e115857, 2019. <https://doi.org/10.1016/j.energy.2019.115857>
- [14] J. Liu, M. Y. Cao, D. Bai, and R. Zhang, “Solar radiation prediction based on random forest of feature-extraction», *IOP Conference Series: Materials Science Engineering*, vol. 658, e012006, 2019. <https://doi.org/10.1088/1757-899X/658/1/012006>
- [15] S. Li, H. Ma, and W. Li, “Typical solar radiation year construction using k-means clustering and discrete-time Markov chain,” *Applied Energy*, vol. 205, pp. 720-731, 2017. <https://doi.org/10.1016/j.apenergy.2017.08.067>
- [16] R. Meenal, and A. I. Selvakumar, “Assessment of SVM, Empirical and ANN based solar radiation prediction models with most influencing input parameters,” *Renewable Energy*, vol. 121, pp. 324-343, 2017. <https://doi.org/10.1016/j.renene.2017.12.005>
- [17] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, “Models for solar radiation prediction based on different measurement sites,” *Measurement*, vol. 63, pp. 346-363, 2015. <https://doi.org/10.1016/j.measurement.2014.11.037>
- [18] E. S. Mostafavi, S. M. Mousavi, and P. Jiao, “Next Generation Prediction Model for Daily Solar Radiation on Horizontal Surface Using a Hybrid Neural Network and Simulated Annealing Method,” *Energy conversion and management*, 2017. <http://agris.fao.org/agris-search/search.do?recordID=US201800045628>
- [19] H. Sharadga, S. Hajimirza, and R. S. Balog, “Time series forecasting of solar power generation for large-scale photovoltaic plants,” *Renewable Energy*, vol. 150, pp. 797-807, 2020. <https://doi.org/10.1016/j.renene.2019.12.131>
- [20] A. Rabehi, M. Guermoui, and D. Lalmi, “Hybrid models for global solar radiation prediction: a case study», *International Journal of Ambient Energy*, vol. 41 (1), pp. 31-40, 2020. <https://doi.org/10.1080/01430750.2018.1443498>
- [21] C. Koo, W. Li, S. H. Cha, and S. Zhang, “A novel estimation approach for the solar radiation potential with its complex spatial pattern via machine-learning techniques,” *Renewable Energy*, vol. 133, pp. 575-592, 2019. <https://doi.org/10.1016/j.renene.2018.10.066>
- [22] S. Hussain, and A. AlAlili, “A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks,” *Applied Energy*, vol. 208, pp. 540-550, 2017. <https://doi.org/10.1016/j.apenergy.2017.09.100>