

Sistema para predicción de la generación en bloques de plantas fotovoltaicas

System for prediction of generation in blocks of photovoltaic plants

Francisco B. Herrera Fernández^I *, Alberto A. Limonte Ruiz^I, Michel Alvarez Morales^I, Jesús G. García Tamayo^{II}

^IUniversidad Central “Marta Abreu” de Las Villas, Villa Clara, Cuba

^{II}Empresa Tecnologías de la Información y la Automática, Villa Clara, Cuba

*Autor de correspondencia: herrera@uclv.edu.cu

Recibido: 14 de agosto de 2023

Aprobado: 16 de octubre 2023

Este documento posee una [licencia Creative Commons Reconocimiento-No Comercial 4.0 internacional](https://creativecommons.org/licenses/by-nc/4.0/) 

RESUMEN/ABSTRACT

La predicción de la generación energética en plantas fotovoltaicas conectadas a un sistema eléctrico es objeto de constante estudio y desarrollo. En objetivo de este trabajo es desarrollar un método para la predicción de la generación a corto plazo. Esta predicción se realiza a partir de la aplicación de programas de predicción para plantas fotovoltaicas desarrollados basados en redes neuronales recurrentes y convolucionales. Como datos de entrada se consideran irradiación solar y temperatura ambiente. Se aplican métodos experimentales, utilizando los datos históricos del comportamiento de estas variables, y aplicando diferentes métodos de preprocesamiento de estos datos y postprocesamiento de las predicciones básicas, lo cual proporciona predicciones con mayor exactitud. Como principal resultado se obtiene un método de predicción, con una mejor exactitud en las horas centrales del día. Se presentan los resultados de las predicciones para un grupo de plantas demostrándose la factibilidad de aplicación del método.

Palabras clave: modelado de plantas fotovoltaicas, predicción de generación.

The prediction of energy generation in photovoltaic plants connected to an electrical system is the subject of constant study and development. The objective of this work is to develop a method for short-term generation prediction. This prediction is made from the application of prediction programs for photovoltaic plants developed based on recurrent and convolutional neural networks. Solar irradiation and ambient temperature are considered as input data. Experimental methods are applied, using historical data on the behavior of these variables, and applying different methods of pre-processing these data and post-processing the basic predictions, which provides predictions with greater accuracy. The main result is a prediction method, with better accuracy in the central hours of the day. The results of the predictions for a group of plants are presented, demonstrating the feasibility of applying the method.

Keywords: modeling of photovoltaic plants, generation forecasting.

INTRODUCCIÓN

En la actualidad una de las más importantes fuentes de generación de energía eléctrica es la energía solar fotovoltaica, clasificada en el campo de las energías limpias y renovables. La dependencia que tiene la energía eléctrica generada en un parque fotovoltaico de las condiciones meteorológicas, y la alta variabilidad de estas, hacen que el problema de la predicción de la energía generada en el mismo sea una labor compleja [1].

Cómo citar este artículo:

Francisco B. Herrera Fernández y otros. Sistema para predicción de la generación en bloques de plantas fotovoltaicas.

Ingeniería Energética. 2023. 44 (3), septiembre/diciembre. ISSN: 1815-5901.

Sitio de la revista: <https://rie.cujae.edu.cu/index.php/RIE/index>

No obstante, está demostrado la necesidad de estimar la producción de energía eléctrica por esta vía dado lo que significa para la correcta planificación y estabilidad del sistema eléctrico. Con ello se logra un aumento del nivel de penetración del sistema fotovoltaico masi como reduce el costo de mantenimiento de los dispositivos auxiliares. Por lo tanto los sistemas de predicción de la generación eólica se fortalecen para contribuir a este objetivo. Además, es una acción de gran importancia y ayuda a los operadores y diseñadores del sistema de energía a modelar y gestionar las plantas solares fotovoltaicas de manera eficiente [2].

Actualmente se trabaja intensamente en la aplicación de herramientas de la Inteligencia Artificial (IA) para el desarrollo de estos tipos de modelos, basado en la demostrada capacidad de estas técnicas en el manejo de información contenida en grandes volúmenes de datos obtenidos de los sistemas objeto de trabajo. Se presentan en la literatura especializada varios métodos de resolver este problema. Algunos usan métodos de predicción indirecto de la potencia fotovoltaica generada, que se basa en realizar la predicción de algunos parámetros meteorológicos (fundamentalmente la irradiación solar) y posteriormente a partir de estas predicciones y mediante un modelo de la planta fotovoltaica que generalmente es un modelo paramétrico, realizan la predicción de la potencia fotovoltaica generada. Se ha trabajado en aplicación de redes de gran memoria a corto plazo (LSTM por sus siglas en inglés) [3], para la predicción de la irradiación solar del día siguiente usando como variables de entrada la temperatura, la velocidad del viento, entre otras.

Se obtiene un error cuadrático medio inferior a 19 %. Otros métodos aplican una metodología espacio-temporal para pronosticar la producción de energía de una planta específica [4, 5]. Obteniéndose un error medio normalizado de un 20 % en promedio. Otros métodos también proporcionan predicciones con similares índices de exactitud. El método basado en ejes de regresión adaptativa [6]. Se ha aplicado un modelo de regresión de vectores de apoyo(SVR por sus siglas en inglés) para predecir la potencia de una planta fotovoltaica para un horizonte de tiempo de corto plazo de 15 minutos hasta 5 horas [7]. Utiliza como entradas mediciones de potencia fotovoltaica y pronóstico de la irradiancia solar. El modelo es capaz de generar buenas predicciones para condiciones de cielo claro y nublado con un valor del RMSE menor de 15%. Se han desarrollado métodos plazo basados en diversas estructuras de redes neuronales artificiales y la lógica difusa, en el cual se usa como variables de entrada la temperatura, el punto de rocío, la velocidad y dirección del viento y la irradiación solar.

En algunos casos con muy buenos índices de exactitud, por ejemplo 10% de error medio absoluto [8], otros con indicadores menos favorables con error medio absoluto de 30 % [9]. En específico existen varias referencias de uso de LSTM en sistemas de predicción de potencia fotovoltaicas, con distintos valores de índices de exactitud; 21% de error medio cuadrático [10]; 11, 8 % de error medio absoluto [11]; con redes neuronales de convolución profunda, 12% de error medio absoluto [12]. Dando continuidad a esta labor reflejada en estas referencias en este trabajo se aplican las técnicas basadas en redes LSTM y redes de convolución con aprendizaje profundo (LSTM y CNN) para realizar la predicción directa de la potencia fotovoltaica generada utilizando solamente un conjunto de mediciones históricas de las plantas analizadas.

Este trabajo se diferencia de los anteriores en la metodología utilizada y en la estructura del modelo predictivo que se propone para realizar la predicción de la potencia fotovoltaica generada, sin necesidad de disponer de predicciones de modelos meteorológicos o de otros modelos de predicción. También se diferencia en la estructura y el tratamiento de los datos históricos recopilados. Además, el método de predicción propuesto permite obtener resultados satisfactorios en cuanto a la calidad de las predicciones realizadas. Todo el trabajo realizado ha estado enmarcado y forma parte del proyecto “Perfeccionamiento del sistema de pronóstico energético para plantas eólicas y fotovoltaicas conectadas el sistema eléctrico nacional”, proyecto perteneciente al Programa Nacional de Ciencia y Técnica Desarrollo Energético Integral y Sostenible.

DESARROLLO

Se desarrolla esta experiencia de predicción a un conjunto de 23 plantas fotovoltaicas, ubicadas en diferentes zonas del país. El método general empleado es:

- Obtención de todos los datos históricos en un periodo de 2 años de las principales variables en todas las plantas en donde se empleará el método de predicción. Variables: Irradiancia, Temperatura ambiente, potencia generada.
- Caracterización y preprocesamiento de los datos históricos en un periodo de 2 años. Variables: Temperatura ambiente, potencia generada. Eliminación de valores anormales y completamiento.
- Aplicación del método LSTM/CNN desarrollado para la predicción de generación fotovoltaica [13]
- Postprocesamiento de los resultados de la predicción.
- Comparación con los métodos actualmente empleados.

A continuación se presentan el desarrollo de estos aspectos.

Caracterización de los datos

Durante el desarrollo de este trabajo, se encontró que la calidad de los datos reales obtenidos de las plantas no era buena, presentando valores faltantes y algunos valores que no tenían sentido, lo que podía afectar negativamente la precisión de la predicción. Para abordar este problema, se implementaron varios métodos de pre procesamiento de datos para mejorar la calidad de los datos y hacerlos más útiles para la predicción.

Mediante el análisis de los datos se evidencia que estos presentan anomalías en sus valores numéricos y faltantes de datos, por lo que se eliminaron los días en que las mediciones se encontraban completamente en cero, y a los datos que presentaban problemas locales (ej. alguna temperatura en cero, algún valor de irradiación por encima de lo normal) se estimaron mediante la función de MatLab `misdata` [14].

Debido a que la irradiación solar (variable principal de entrada al modelo) es nula en horas de la noche, al igual que la variable potencia activa (salida del modelo), se decidió tomar doce horas del día en las cuales estas variables presentaban una dinámica no nula. Se corrobora que la principal correlación de la potencia generada es con irradiación y con temperatura de los módulos, lo cual se aprecia en la figura 1, [14].

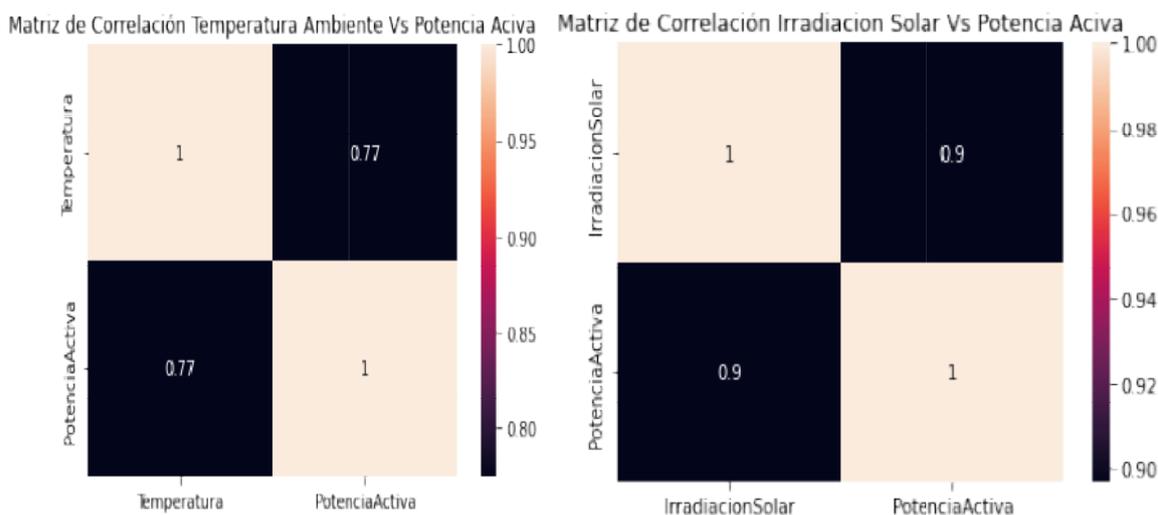


Fig. 1. Matrices de correlación entre las variables analizadas [14]

Una vez en el `SystemIdentificationToolbox` del Matlab los datos fueron preprocesados, removiendo tendencias y eliminando medias, además fueron divididos en dos conjuntos de datos; 75% para identificación y 25% para la validación del modelo.

Los datos fueron organizados de la siguiente forma atendiendo a la estructura del modelo:

- Entradas:
 1. Irradiancia solar [W/m²],
 2. temperatura ambiente [°C],
 3. temperatura de los paneles fotovoltaicos [°C],
 4. potencia generada [MW].
- Salida:
 1. Predicción de la potencia fotovoltaica generada para el día siguiente.

La transformación de los datos primarios se realiza en varias etapas de preprocesamiento:

- Eliminación de los atributos que no aportan información relevante
- Sustitución de valores inconsistentes (nulos y no numéricos) por el valor promedio del atributo al que correspondiera
- Normalización de los valores

Se emplearon tres métodos para este procesamiento:

Reemplazar los datos faltantes por los promedios de los días anteriores. Calcular los promedios en cada una de las horas, que luego se utilizan para reemplazar los valores ausentes en el conjunto de datos.

Como inicio al proceso de explotación de los modelos seleccionados se tuvo que conformar un conjunto de entrenamiento, para el cual se tomaron datos de la fecha, la irradiación solar, la temperatura ambiente, la temperatura del módulo y la potencia real para cada una de las horas necesarias que van desde las 7:00:00 am hasta las 18:00:00 pm. De esta forma, se creó un conjunto de entrenamiento para cada uno de los parques seleccionados, los cuales se eligieron teniendo en cuenta la calidad de los datos dados en cuanto a cantidad de valores faltantes.

Una vez conformados los conjuntos de entrenamiento se entrenaron modelos por cada uno de los parques y se guardaron dichos modelos una vez entrenados para luego ser usados en la realización de las predicciones. Como principal conclusión de la implementación de los programas de preprocesamiento se obtiene una significativa mejoría de la calidad de los datos y la precisión de la predicción. Estos resultados demuestran la importancia de la calidad de los datos y la implementación adecuada de métodos de pre procesamiento de datos en proyectos de aprendizaje automático y análisis de datos.

Modelo de predicción

Los principales parámetros de cada modelo implementado se resume en las tablas 1 y 2, respectivamente

Tabla 1. Parámetros del modelo LSTM

Parámetros del modelo	Descripción
Número de capas	4
Neuronas en la primera capa	64
Neuronas en la segunda capa	32
Neuronas en la tercera capa	19
Neuronas capa de salida	1
Función de activación capa salida	Lineal
Tasa de aprendizaje inicial	0.01
Función objetivo	mse
Optimizador	<i>Estimación adaptativa del momento (Adam)</i>
Épocas de entrenamiento	30
Tamaño del lote	256

Tabla 2. Parámetros del modelo CNN

Parámetros del modelo	Descripción
Capa Conv1D	Filters =50, kernel_size = 20
Capa Maxpooling1D	Pool_size = 2
Capa Dense	50 neuronas
Capa Dense	1 neurona
Función de activación	RELU
Tasa de aprendizaje inicial	0.01
Función objetivo	mse
Optimizador	<i>Estimación adaptativa del momento (Adam)</i>
Épocas de entrenamiento	200
Tamaño del lote	256

Evaluación del programa de predicción

La evaluación del programa de predicción se comenzó en el mes de febrero 2023 aplicando el programa a la predicción de generación de varias plantas (3 plantas inicialmente y luego se extendió a 23 plantas). El programa toma los datos 14 días anteriores a la fecha actual. Con estos datos, preprocesados previamente, se forma un conjunto de datos cuyas columnas son: fecha, irradiación, temperatura del módulo, temperatura ambiente y la potencia generada, en cada una de las horas usadas (de las 07:00:00 a las 18:00:00). Una vez pasados los datos, al modelo del parque correspondiente, éste devuelve las predicciones de los 14 días posteriores, de las cuales solo se toman las del día deseado (día siguiente al día actual), esto se repite para cada uno de los parques de los cuales se cuenta con modelos y luego se pasan dichas predicciones a un documento csv.

Se confeccionó un programa para la lectura de este fichero Excel, el cálculo del error medio porcentual de estimación en cada día y por cada hora, usando la ecuación (1):

$$Error_{(\%)} = \frac{(Potencia_{real} - Potencia_{pronosticada})}{Potencia_{real}} \quad (1)$$

Se calcula además el error medio en el día, promediando el error (valor absoluto) obtenido en cada una de las horas del día.

RESULTADOS

En las siguientes figuras se muestran los resultados obtenidos para la predicción de la generación para los primeros 20 días del mes de agosto de este año 2023, para tres de las 23 plantas analizadas. En la figura 2, se compara la predicción básica que de la potencia generada realiza el modelo, comparada con los valores reales, para las distintas horas prácticas o útiles, o sea de 7.00 a las 18.00 horas. Como se aprecia existe una diferencia notable entre la predicción básica y los valores reales, lo cual justifica el necesario postprocesamiento de estas predicciones. Como criterio para la estimación de la corrección se usó el propio error obteniéndose la ley de comportamiento del mismo, la cual se muestra en la figura 3.

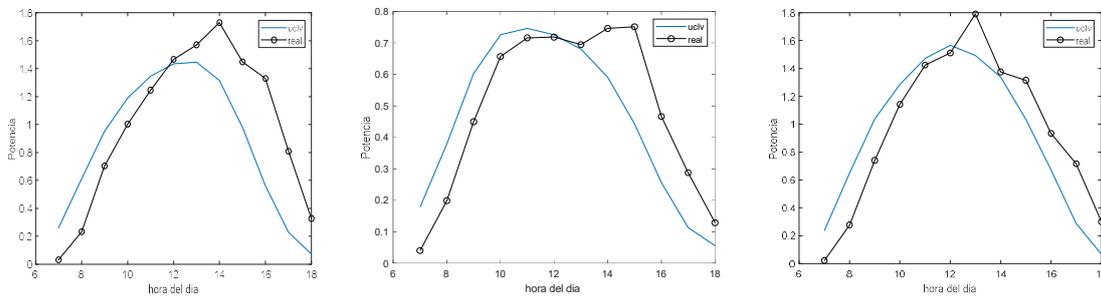


Fig. 2. Resultado de la predicción básica comparado con los valores reales para tres plantas

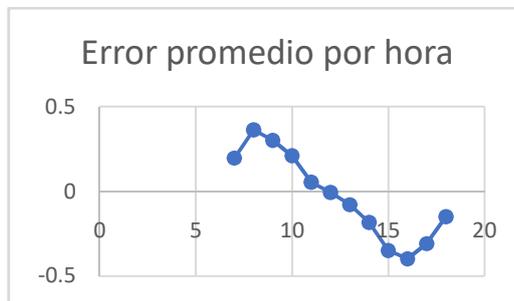


Fig. 3. Comportamiento del error promedio del predicción por hora útil en 20 días

Considerando el comportamiento del error de predicción como variable de ajuste en el modelo, retroalimentando la misma, en el modelo de convolución se obtiene una nueva predicción, o predicción ajustada, la cual se muestra en la figura 4, para estas mismas plantas.

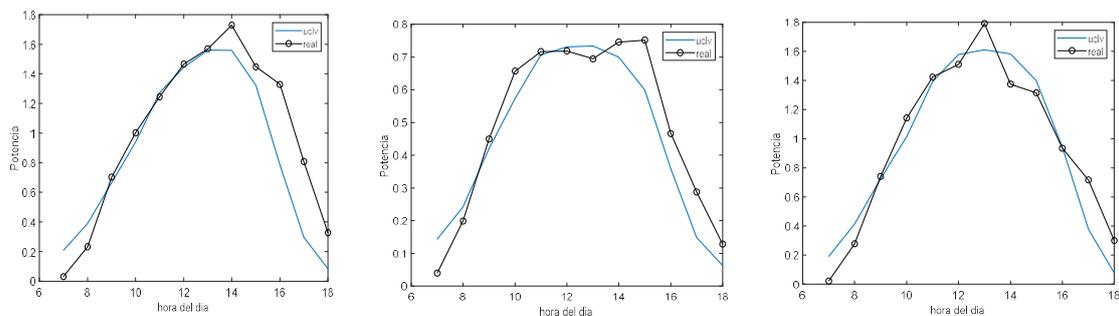


Fig. 4. Resultado de la predicción ajustada comparado con los valores reales para tres plantas

Se puede apreciar la significativa mejoría en la predicción. En la figura 5 se muestra como es el error de predicción (en unidades de potencia generada) para esta predicción ajustada. Se aprecia que se cumple que para la primera mitad del día la predicción es superior al valor real, invirtiéndose esta relación en la segunda mitad del día. En todos los casos se aprecia que el error está por debajo de 0.1 MW para prácticamente todo el día. Si consideramos que 1 MW es un valor estándar para la generación de cualquiera de estas plantas, entonces el error relativo obtenido está en el orden del 10 % lo cual es comparable y en algunos casos mejor que los métodos presentados en la literatura. Particularmente se analiza continuación el error relativo que se presenta en las horas centrales del día útil, o sea entre las 10.00 y las 14.00 horas. En la figura 6 se muestra el error para estas horas. Aplicando el programa y método de predicción en el resto de las plantas analizadas se obtienen similares resultados, por lo que se concluye la validez y generalidad del método

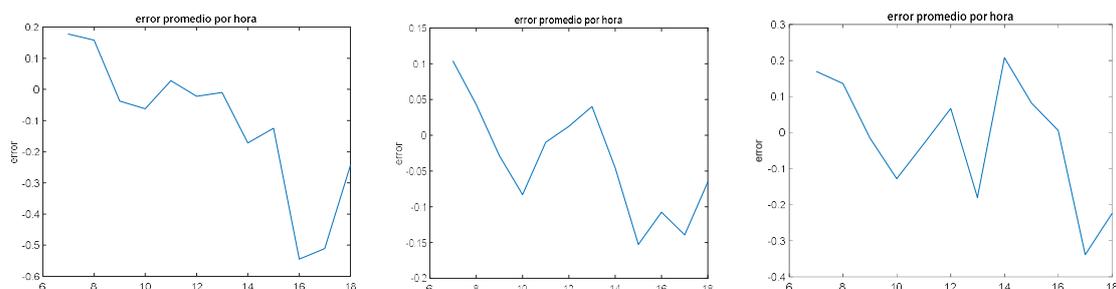


Fig. 5. Error de predicción, valor absoluto de potencia, para la predicción ajustada

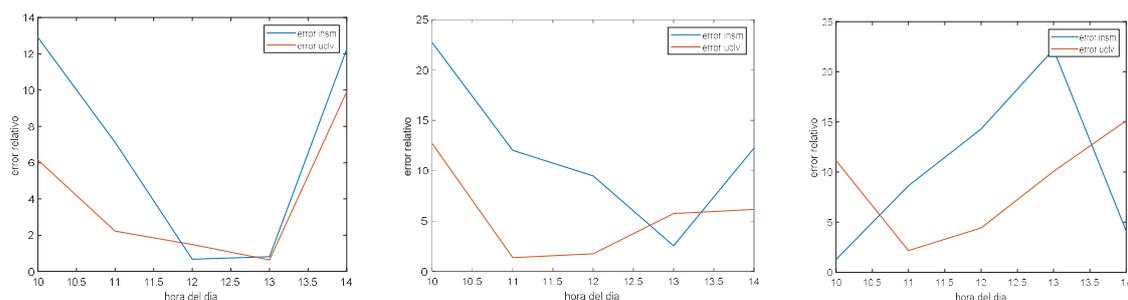


Fig. 6. Comparación nuevo método con método actual de predicción para las horas centrales del día Rojo uciv método nuevo / azul insm método actual

CONCLUSIONES

El método desarrollado, hasta donde está ajustado en la parte del preprocesamiento de los datos para el aprendizaje y el postprocesamiento tiene un comportamiento en general similar al actualmente empleado, con una marcada mejoría en las horas centrales del día (las de mayor generación de potencia). Esto motiva la conclusión que el método es factible de aplicarse a gran escala como un segundo método de predicción en plantas fotovoltaicas del país. Estos resultados de la evaluación indica además la necesidad de continuar mejorando el postprocesamiento de la predicción.

REFERENCIAS

- [1]. Jiménez, A. M. “Modelos de predicción a corto plazo de la generación eléctrica en instalaciones fotovoltaicas”. Tesis de doctorado. Universidad de la Rioja. Logroño, España, 2014. Disponible en: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj7k97ggJ6CAxUOIWoFHedJAzUQFnoECAGQAO&url=https%3A%2F%2Fdia.net.unirioja.es%2Fdescarga%2Ftesis%2F44089.pdf&usg=AOvVaw3j0J8bShxKWRYIk StNHvY&opi=89978449>
- [2]. Ahmed, R., *et al.* “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization”. Renewable and Sustainable Energy Reviews [en línea]. Mayo 2020. vol. 124, e109792. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S1364032120300885>
- [3]. Qing, Xiangyun y Niu, Yugang. “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM”. Energy [en línea]. Abril 2018, vol. 148, p. 461-468. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0360544218302056>

- [4]. Lima, Francisco J. L., *et al.* "Forecast for surface solar irradiance at the Brazilian Northeastern región using NWP model and artificial neural networks". *Renewable Energy* [en línea]. Marzo 2016, vol. 87, p. 807-818. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0960148115304249>
- [5]. Agoua, Xwegnon Ghislain, *et al.* "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production". *IEEE Transactions on Sustainable Energy* [en línea]. Abril 2018, vol. 9, n. 2, p. 538-546. ISSN 1949-3037. Disponible en: <https://ieeexplore.ieee.org/document/8023858>
- [6]. Massidda, Luca y Marrocu, Marino. "Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany". *Solar Energy* [en línea]. Abril 2017, vol. 146, p. 141-149. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0038092X17301007>
- [7]. Wolff, Björn, *et al.* "Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data". *Solar Energy*. Octubre 2016, vol. 135, p. 197-208. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0038092X16301682>
- [8]. Raza, Muhammad Qamar, *et al.* "An Ensemble Framework for Day-Ahead Forecast of PV Output Power in Smart Grids". *IEEE Transactions on Industrial Informatics*. Agosto 2019, vol. 15, n. 8, p. 4624-4634. ISSN 1941-0050. Disponible en: <https://ieeexplore.ieee.org/document/8542710>
- [9]. Sivaneasan, B., *et al.* "Solar Forecasting using ANN with Fuzzy Logic Pre-processing". *Energy Procedia*. Diciembre 2017, vol. 143, p. 727-732. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1876610217365177>
- [10]. Beristáin Jimenez, J. A. " Inversores bidireccionales con aislamiento en alta frecuencia para aplicaciones de energías renovables". Tesis Doctoral. Universidad Politécnica de Cataluña. España [en línea]. 17 diciembre 2021. Disponible en: <https://www.tdx.cat/handle/10803/6336?show=full&locale-attribute=ca>
- [11]. Wen, Lulu, *et al.* "Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting". *Energy*. Marzo 2019, vol. 171, p. 1053-1065. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0360544219300775>
- [12]. Zhang, Rui, *et al.* "Forecast of Solar Energy Production - A Deep Learning Approach". En: 2018 IEEE International Conference on Big Knowledge (ICBK). Singapore: IEEE. Noviembre 2018. p. 73-82. ISBN 978-1-5386-9125-0. Disponible en: <https://ieeexplore.ieee.org/abstract/document/8588777>
- [13]. Herrera Casanova, Reinier, *et al.* "Propuesta de modelo predictivo inteligente para una planta fotovoltaica". *Revista Cubana de Ciencias Informáticas*. 2020, vol. 16, n. 1, p. 144-162. ISSN 2227-1899. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992022000100144
- [14]. Garcia Tamayo, Jesus y Herrera Fernández, Francisco. "Modelado de un parque solar fotovoltaico de empleo industrial". *Ingeniería Energética*. 2023, vol. 44, n.1, p. 1-7. ISSN 1815-5901. Disponible en: <http://scielo.sld.cu/pdf/rie/v44n1/1815-5901-rie-44-01-125.pdf>

CONFLICTO DE INTERESES

Los autores declaran que no hay conflictos de intereses.

CONTRIBUCIÓN DE LOS AUTORES

Francisco Herrera Fernández: <https://orcid.org/0000-0002-0774-0752>

Revisión del estado del arte, diseño de la investigación, recolección de datos y procesamiento, simulación, análisis de los resultados ,redacción del artículo y revisión crítica del contenido hasta su aprobación final.

Alberto Limonte Ruiz: <https://orcid.org/0000-0002-7272-1216>

Revisión del estado del arte, análisis de los resultados y la revisión crítica de su contenido y la aprobación final.

Michel Alvarez Morales: <https://orcid.org/0009-0007-2662-7433>

Revisión del estado del arte, recolección de datos y procesamiento, programación y corrida del software, simulación, análisis de los resultados , redacción del borrador del artículo y la aprobación final.

Jesús G. García Tamayo: <https://orcid.org/0000-0002-8158-6984>

Revisión del estado del arte, recolección de datos y procesamiento, programación y corrida del software, simulación, análisis de los resultados , redacción del borrador del artículo y la aprobación final.