

ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

<https://dx.doi.org/10.14482/inde.43.01.456.089>

E-solar: una herramienta para la evaluación del recurso solar basada en una arquitectura *big data* sobre un ambiente PySpark

E-solar: a tool for solar resource assessment based on a Big Data architecture in a PySpark environment

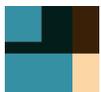
LUIS EDUARDO ORDÓÑEZ PALACIOS*
VÍCTOR BUCHELI GUERRERO**
EDUARDO CAICEDO BRAVO***

* Investigador, Universidad del Valle (Colombia), Escuela de Ingeniería de Sistemas y Computación (EISC). Doctor en Ingeniería. Orcid-ID: <https://orcid.org/0000-0001-5154-9472>. luis.ordonez.palacios@correounivalle.edu.co

** Profesor titular, Universidad del Valle (Colombia), Escuela de Ingeniería de Sistemas y Computación (EISC). Doctor en Ingeniería. Orcid-ID: <https://orcid.org/0000-0002-0885-8699>. victor.bucheli@correounivalle.edu.co

*** Profesor titular, Universidad del Valle (Colombia), Escuela de Ingeniería Eléctrica y Electrónica (EIEE). Doctor en Ingeniería. Orcid-ID: <https://orcid.org/0000-0003-0727-2917>. eduardo.caicedo@correounivalle.edu.co

Correspondencia: Luis Eduardo Ordóñez Palacios. Cel. 311 374 1912. Dirección: Cali, Carrera 47 n.° 13B-23.



Resumen

Con el tiempo, diversos investigadores han creado modelos matemáticos, estadísticos y predictivos para evaluar el recurso solar. Sin embargo, su implementación en herramientas técnicas limita su utilización por usuarios no técnicos. Además, el procesamiento de datos para estimar la radiación solar suele requerir hardware potente. Este estudio presenta una herramienta basada en Big data que utiliza archivos planos e imágenes de satélite para estimar la radiación solar en Colombia. Se desarrolló un modelo con técnicas de aprendizaje automático y varios lenguajes de programación. Se ejecuta en MapR, una distribución del ecosistema Hadoop con un amplio conjunto de capacidades big data y emplea la API de PySpark para procesar datos en paralelo en un clúster de computadoras. La herramienta E-solar implementada en un servidor web fue evaluada por profesionales del sector energético. Se analizó la usabilidad, se verificó la conformidad con estándares de programación recientes y se identificaron perfiles de usuarios interesados. Los datos de radiación solar generados por la herramienta son fundamentales para proyectos solares. Además, la herramienta proporciona apoyo a investigadores y organizaciones; y facilita la toma de decisiones en la implementación de sistemas fotovoltaicos al ofrecer información relevante sobre el comportamiento del recurso solar en Colombia.

Palabras clave: aprendizaje automático, big data, MapR, PySpark, radiación solar.

Abstract

Over time, diverse researchers have created mathematical, statistical, and predictive models to evaluate solar resources. However, their implementation in technical tools restricts their usability for non-technical users. Additionally, data processing to estimate solar radiation often necessitates powerful hardware. This study introduces a Big Data based tool that employs flat files and satellite images to estimate solar radiation in Colombia. A model was developed using machine learning techniques and various programming languages. It operates within MapR, a distribution of the Hadoop ecosystem with an extensive array of Big Data capabilities and utilizes the PySpark API for parallel data processing within a computer cluster. The E-Solar tool, deployed on a web server, underwent assessment by professionals within the energy sector. Usability was analyzed, compliance with recent programming standards was confirmed, and profiles of interested users were identified. The solar radiation data generated by the tool are pivotal for solar projects. Furthermore, the tool lends support to researchers and organizations in decision-making for the implementation of photovoltaic systems, as it offers pertinent information regarding the behavior of solar resources in Colombia.

Keywords: Big Data, machine learning, MapR, PySpark, solar radiation.

INTRODUCCIÓN

La generación de energía eléctrica actual está orientada al uso de fuentes de energía no convencionales. En este sentido, se están implementando diversos proyectos que utilizan la energía solar por considerarse una fuente de energía limpia e inagotable. Asimismo, se están haciendo grandes esfuerzos para mitigar el cambio climático, como se observa en el estudio de Zheng et al. [1]. Entre otras razones, es por esto que los investigadores están diseñando modelos matemáticos, estadísticos y predictivos del recurso solar para asistir a los responsables de la toma de decisiones en la planificación e implementación de sistemas fotovoltaicos.

En esta investigación se consideró el trabajo de Ordóñez Palacios et al. [2], basado en la implementación de un modelo de Aprendizaje Automático para la estimación de la radiación solar en Colombia a partir de imágenes obtenidas del satélite GOES-13 [3]. En el estudio se evaluaron 4 técnicas de aprendizaje automático, de las cuales el algoritmo Random Forest presentó el rendimiento más alto según las métricas R^2 y RMSE, con valores de 0,82 y 107,25, respectivamente. Por consiguiente, se desarrolló el dashboard E-solar, que incluye el modelo Random Forest, con el objetivo de ponerla a disposición de los investigadores alrededor de la evaluación del recurso solar en Colombia.

La herramienta web E-solar se implementó sobre una arquitectura *big data*, que consiste en una infraestructura de datos a gran escala, refiriéndose al diseño y estructura de los sistemas informáticos destinados a recolectar, almacenar, procesar y analizar grandes volúmenes de datos, a menudo en tiempo real o con corta diferencia del tiempo real. Estos datos pueden ser estructurados (como bases de datos), semiestructurados (como XML o JSON) o no estructurados (como textos, imágenes o videos).

Esta solución satisface los requerimientos de gestión de datos heterogéneos, almacenamiento y procesamiento de grandes volúmenes de información, provenientes de archivos planos e imágenes satelitales. Las funcionalidades de E-solar están soportadas por el *framework* de código abierto PySpark [4] como interfaz de Python para Apache Spark [5]; esto posibilita el procesamiento de datos a gran escala y lo convierte en una herramienta esencial para el ecosistema de *big data*, permitiendo la computación en paralelo sobre un clúster de 5 computadores. La herramienta está desarrollada utilizando Python como lenguaje principal, Django [6] como *framework* web y JavaScript [7] para la interactividad del lado del cliente, el código fuente está alojado en el repositorio de GitHub en [8].

El desarrollo del *dashboard* E-solar se constituye en una alternativa del proyecto Power Data Access Viewer de la NASA [9] para la evaluación del recurso solar en Colombia. Aunque es importante resaltar que el visor de datos de la NASA proporciona



datos solares y meteorológicos de la investigación de la NASA para apoyar la energía renovable, la eficiencia energética de los edificios, entre otras aplicaciones, a escala mundial. En contraste, E-solar únicamente permite la evaluación del recurso solar a partir de datos contenidos en archivos planos y la generación de datos de radiación solar en Colombia gracias al procesamiento de imágenes satelitales entre 2012 y 2017.

E-solar incluye los siguientes algoritmos de regresión para evaluar conjuntos de datos cargados al sistema: AdaBoost Regressor, Árboles de Regresión, Random Forest para Regresión (RF), Gradient Boosting Regressor, Redes Neuronales Artificiales, Regresión con soporte vectorial (SVR), Regresión Lineal Múltiple, Regresión Polinomial y XGBoost Regressor. Asimismo, el *dashboard* utiliza el algoritmo Random Forest para la generación de datos de radiación solar en Colombia. Toda la información relacionada con las técnicas de aprendizaje automático mencionadas anteriormente se encuentra en la documentación registrada en los sitios web de cada proyecto [10], [11].

El portal de consulta y descarga de datos hidrometeorológicos [12] del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), ofrece un banco de datos de información recopilada por estaciones de monitoreo ubicadas en diversas regiones del país. Sin embargo, el número de laboratorios dedicados a la medición de variables meteorológicas es limitado. Por esta razón, el *dashboard* E-solar brinda apoyo en la generación de datos de radiación solar en cualquier región del territorio colombiano.

En un contexto de creciente interés en fuentes de energía sostenible y reducción del cambio climático se presenta el proyecto E-solar. Este utiliza el modelo de Aprendizaje Automático Random Forest para evaluar y generar datos de radiación solar en Colombia. Se espera que el *dashboard* E-solar sea una herramienta valiosa para investigadores y tomadores de decisiones, ofreciendo datos precisos sobre el recurso solar en el país, diferenciándose de otras fuentes de datos limitadas a la medición de variables meteorológicas.

Este trabajo contribuye significativamente al campo de la investigación en energías renovables, específicamente en la evaluación del recurso solar, mediante la integración de tecnologías de *big data* y aprendizaje automático. La investigación aborda dos desafíos importantes: la accesibilidad para usuarios no técnicos y la demanda de recursos computacionales en el procesamiento de datos solares. E-solar representa un avance importante en el uso de ambientes de *big data* basados en MapR y PySpark [4] para procesar datos heterogéneos, incluyendo archivos planos e imágenes de satélite. Este enfoque mejora la eficiencia computacional y facilita el acceso a la información sobre radiación solar para diversos usuarios. Además, la usabilidad y adherencia a estándares de desarrollo modernos aumenta la relevancia y el impacto

potencial en la toma de decisiones para proyectos de energía solar en Colombia, impulsando el desarrollo sostenible y la transición energética.

Este documento está integrado por las siguientes secciones: trabajos relacionados, la metodología, una descripción detallada sobre el desarrollo del proyecto, los resultados obtenidos sobre los casos de estudio aplicados en Colombia, y finalmente, las conclusiones.

TRABAJOS RELACIONADOS

Es común encontrar el uso de *big data* en proyectos de diversas áreas del conocimiento, como en el estudio de Wu et al. [13], en el que se desarrolla un sistema que integra operaciones de maquinaria agrícola con *big data* y presenta un estudio de caso con cosechadores de trigo para demostrar su potencial.

Existe suficiente literatura sobre investigaciones relacionadas con la predicción de radiación solar, sin embargo, muchos de los trabajos llegan hasta el modelado y las simulaciones, pero no se integran en aplicaciones para uso del público en general. El trabajo de Rodríguez et al. [14] utiliza redes neuronales artificiales para predecir la generación de energía solar fotovoltaica. La herramienta fue desarrollada en el software MATLAB y utilizaron el error cuadrático medio para la validación de resultados, con un error de entre 0,5 y 9 %.

El trabajo de Boutahir et al. [15] desarrolla un conjunto de modelos de aprendizaje profundo a través de algoritmos que verifican la importancia en la selección de parámetros meteorológicos en una ubicación específica de la región de Errachidia, en Marruecos, para predecir los datos de radiación solar directa. Los hallazgos demostraron que este enfoque desempeña un papel crucial en la predicción precisa de la radiación solar en comparación con los datos disponibles. Por otro lado, la investigación de Tehrani et al. [16] utilizó un enfoque basado en datos urbanos para entrenar un modelo de redes neuronales artificiales para discernir relaciones complejas y no lineales entre variables independientes y dependientes, lo que permitió realizar predicciones de radiación solar en ciudades urbanas con un error cuadrático medio de 0,01 y un valor de R2 cuadrado del 85 %.

En el estudio de Cuesta et al. [17] se comparan dos arquitecturas de redes neuronales recurrentes enfocadas en la predicción de radiación solar y temperatura. Los resultados presentan coeficientes de correlación superiores a 0,9 en la etapa de validación, superiores a 0,8 en la etapa de predicción y el error cuadrático medio muestra valores inferiores a 0,02 kW de radiación solar y 2 °C de temperatura ambiente. De igual manera, la investigación de Cárdenas et al. [18] utiliza la arquitectura neuronal,

Deep Belief Network (DBN), para la predicción de radiación solar. Las simulaciones se realizaron en la herramienta de programación Visual Studio C#. Se obtuvieron errores cercanos al 2 % frente a errores cercanos al 5% de error cuadrático medio, obtenido por trabajos que utilizan técnicas convencionales de redes neuronales como el perceptrón multicapa MLP.

Por su parte, la investigación de Vinod et al. [19] presenta el modelado y la simulación de un módulo fotovoltaico basado en un ambiente de Matlab/Simulink. El resultado del estudio presentó un porcentaje relativo de error del 1,65 %. Otros estudios, como el de Pendem y Mikkili [20], realizaron modelado, simulación y análisis de rendimiento de configuraciones de paneles solares fotovoltaicos en MATLAB/Simulink. En el trabajo de Abdullahi et al. [21] prepararon el modelado del módulo fotovoltaico monocristalino en un ambiente Matlab/Simulink. En este estudio, los resultados demostraron que el módulo es capaz de generar 17,75 W/m² con una eficiencia del 7 % y 138 W/m² con una eficiencia del 8 % en edificios de oficinas interiores y exteriores, respectivamente.

E-solar se fundamenta en investigaciones que abordan la evaluación del recurso solar y el uso de tecnologías avanzadas, como el trabajo de Ordóñez Palacios et al. [2]. Otros estudios han explorado el uso de técnicas, como redes neuronales artificiales y aprendizaje profundo, para predecir la radiación solar con alta precisión. Trabajos como los de Rodríguez et al. [14], Boutahir et al. [15], Tehrani et al. [16] y Cuesta et al. [17] han demostrado la eficacia de estas técnicas, alcanzando errores cuadráticos medios bajos y coeficientes de correlación elevados. Además, investigaciones como las de Cárdenas et al. [18] y Vinod et al. [19] han profundizado en el modelado y simulación de sistemas fotovoltaicos, proporcionando perspectivas valiosas sobre el rendimiento y la eficiencia. Sin embargo, muchos de estos estudios se limitan a modelos y simulaciones sin ofrecer herramientas accesibles al público general. E-solar se distingue por integrar técnicas estudiadas en diversos trabajos con tecnologías de *big data*, como se ve en el trabajo de Wu et al. [13], para crear una herramienta que predice con precisión y hace que los datos sean accesibles y útiles para una amplia gama de usuarios.

METODOLOGÍA

El desarrollo de la herramienta E-solar se proyectó para ejecutarse sobre una arquitectura *big data* con el objetivo de dar solución a los requerimientos de almacenamiento y procesamiento masivo de datos heterogéneos. Posteriormente, se construyó sobre el lenguaje de programación Python, el *framework* Django [6] y algunas librerías de JavaScript. La tabla 1 presenta las fases del *dashboard* y una breve descripción de cada una.

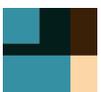


TABLA 1. FASES DEL PROYECTO

Fases del dashboard	Descripción
Fase 1. Planificación del sistema	Aunque existen herramientas para la estimación de radiación solar, E-solar se planificó con una alternativa para evaluar el recurso solar en el territorio colombiano.
Fase 2. Implementación de la arquitectura big data	De acuerdo con la investigación de Ordóñez Palacios et al. [2], el consumo de recursos de hardware por el modelo predictivo es alto. Adicionalmente, el modelo utiliza datos heterogéneos. Estas razones conllevan a la instalación y configuración de MapR sobre un clúster de 5 computadores.
Fase 3. Extracción, transformación y carga de datos	Los datos son gestionados por Apache Hive [22], PostgreSQL y SQLite. Asimismo, la información referente a archivos planos e imágenes de satélite es administrada por el sistema de archivos de Hadoop (HDFS).
Fase 4. Análisis de información	La información es analizada por diferentes algoritmos de aprendizaje automático. Sin embargo, la predicción de datos en cualquier lugar de Colombia está basada en la técnica de regresión Random Forest.
Fase 5. Comunicación de la información	Los resultados del análisis de datos, de las estimaciones de radiación solar y de la evaluación de datos meteorológicos es presentada en diagramas de dispersión, de barras y de líneas para una mejor comprensión por parte del usuario.
Fase 6. Evaluación de la herramienta E-solar	La herramienta E-solar pasó por un proceso de validación realizado por un equipo de 30 profesionales de la industria energética y expertos en ciencias de la computación. Además, recibió evaluaciones por parte de profesionales de reconocidas empresas del sector energético en Colombia, entre las que se incluyen CELSIA, EMCALI, EPM y otras compañías líderes.

Fuente: elaboración propia.

Este trabajo procura dar respuesta a preguntas como: ¿La construcción de una arquitectura big data es necesaria para un proyecto de este tipo?; ¿la herramienta es útil para investigadores y empresarios del sector eléctrico?; ¿cómo desarrollar una herramienta que aproveche el rendimiento de un clúster de computadoras?; ¿cuáles algoritmos de aprendizaje automático se podrían implementar en este proyecto?; ¿qué tipo de visualizaciones son las más adecuadas para presentar la información procedente del análisis de datos? Todos estos interrogantes se resolverán a lo largo del documento.

DESARROLLO DEL PROYECTO

Modelo técnico de la arquitectura *big data*

Considerando el uso de datos heterogéneos para la evaluación del recurso solar, el almacenamiento y el procesamiento de datos masivos de información, se construyó una arquitectura *big data* para dar soporte a la herramienta E-solar. El clúster se implementó sobre 5 computadores con sistema operativo Linux Ubuntu Server 20.04. Se instaló y se configuró la distribución MapR versión 7.2.0 y el paquete del ecosistema MapR (MEP) versión 9.1.0. Finalmente, y con el objetivo de posibilitar la computación en paralelo, se utilizó el *framework* PySpark [4].

Los equipos están conectados a través de una red cableada de datos. Cada máquina está equipada con un procesador Core i7 de 8 núcleos, un disco duro de estado sólido de 500 Gb para el sistema operativo y una memoria RAM de 64 Gb. cuatro de los cinco equipos incluyen un disco duro de estado sólido adicional de 1 Tb para el almacenamiento de datos del clúster. Uno de los equipos está configurado como Web Manager para la gestión de los nodos y los otros cuatro computadores conforman el clúster, donde uno de ellos es el nodo maestro y los demás están configurados como esclavos. En este sentido, el clúster dispone de un procesamiento de datos conformado por 32 núcleos físicos y 32 núcleos lógicos, una memoria RAM de 256 Gb y una capacidad de almacenamiento de 4 Tb.

El proceso de instalación y configuración de MapR está basado en la documentación de HPE Ezmeral Data Fabric [23]. En el sitio web se encuentra todo lo relacionado con la planificación del clúster, la instalación de componentes básicos y demás instrucciones de configuración de clientes, servicios y del sistema de control. Del paquete del ecosistema MapR se instalaron las herramientas: Apache Drill, Hue, Apache Hive, Hadoop HDFS, Spark y el Webserver para el monitoreo del clúster.

Para la gestión de los datos de la herramienta E-solar se utilizó el *framework* Apache Drill [24], el cual admite aplicaciones distribuidas de uso intensivo de datos para el análisis interactivo de conjuntos de datos a gran escala. Hue [25] fue utilizado como asistente SQL de código abierto para consultar bases de datos y almacenes de datos. Asimismo, se utilizó la infraestructura de almacenamiento de datos construida sobre Hadoop, Apache Hive [22], para proporcionar agrupación, consulta y análisis de datos.

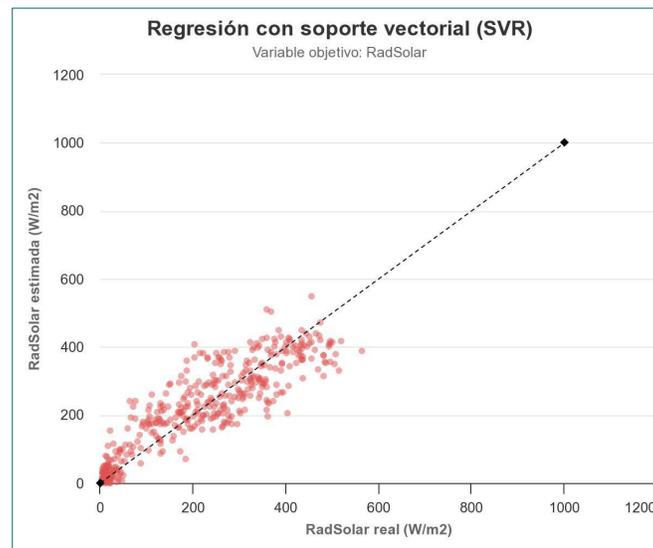
En el caso del almacenamiento de los archivos planos y las imágenes en el clúster, se utilizó el sistema de archivos de Hadoop (HDFS). Por su parte, Apache Spark [5] se utilizó como motor de análisis unificado para el procesamiento de datos a gran escala y posibilitó el procesamiento distribuido de los datos. Por último, el Webserver se

utiliza para la gestión del control del sistema; desde allí se monitorean los servicios en ejecución, los nodos que integran el clúster y el uso de los recursos disponibles.

Desarrollo de la herramienta E-solar

El *dashboard* E-solar fue desarrollado en el lenguaje de programación Python, el *framework* Django [6] y algunas librerías adicionales de JavaScript. Las visualizaciones se desarrollaron con la librería de JavaScript, Highcharts [26]. La herramienta está basada en los gestores de bases de datos: Apache Hive, PostgreSQL y SQLite. Permite la gestión de usuarios, la gestión de archivos planos, el análisis de datos mediante algoritmos de aprendizaje automático, la predicción de radiación solar en cualquier lugar del territorio colombiano, la evaluación de recursos meteorológicos, la visualización de los resultados y la gestión histórica de los datos generados.

E-solar permite subir archivos de texto, archivos csv y archivos de excel que contengan únicamente datos numéricos. El módulo de análisis de datos utiliza los archivos cargados al sistema para realizar predicciones. El usuario debe seleccionar los predictores y la variable objetivo contenidos en el archivo. Finalmente, se selecciona uno de los algoritmos de aprendizaje automático para regresión: AdaBoost Regresor, Árboles de Regresión, Random Forest para Regresión (RF), Gradient Boosting Regresor, Redes Neuronales Artificiales, Regresión con soporte vectorial (SVR), Regresión Lineal Múltiple, Regresión Polinomial y XGBoost Regresor. Los resultados se presentan en un diagrama de dispersión, como se ve en la figura 1, y se almacenan para llevar un control histórico de los análisis realizados.



Fuente: elaboración propia.

FIGURA 1. DIAGRAMA DE DISPERSIÓN CON LOS RESULTADOS DEL ANÁLISIS DE DATOS

El módulo de predicción de radiación solar en cualquier lugar del territorio colombiano utiliza imágenes del satélite GOES-13 [3] de los años entre 2012 y 2017. Se requiere el nombre del lugar, la latitud y la longitud de la ubicación en Colombia para realizar el proceso. El proceso está desarrollado con programación multihilo para que se realice en segundo plano, dado que requiere el uso de todos los recursos de procesamiento del clúster y puede tardar unos segundos. Mientras tanto, el usuario puede utilizar otras funcionalidades del sistema.

Los conjuntos de datos que se suben al sistema y los conjuntos de datos que son generados por E-solar se pueden gestionar desde el módulo de gestión de archivos. Los archivos se pueden abrir en cuadrículas que permiten la búsqueda de información y algunas opciones de visualización. También se pueden descargar o eliminar según las necesidades del usuario.

El módulo de recursos permite evaluar la información meteorológica contenida en los archivos que se suben al sistema. El *dashboard* presenta gráficas de líneas con promedios mensuales de temperatura y de radiación solar. Diagramas de barras horizontales con promedios mensuales de lluvia y humedad. Asimismo, presenta un diagrama compuesto de líneas y barras verticales con promedios mensuales de velocidad y dirección del viento. Los ejemplos mencionados se pueden observar en las figuras 2, 3, 4 y 5.



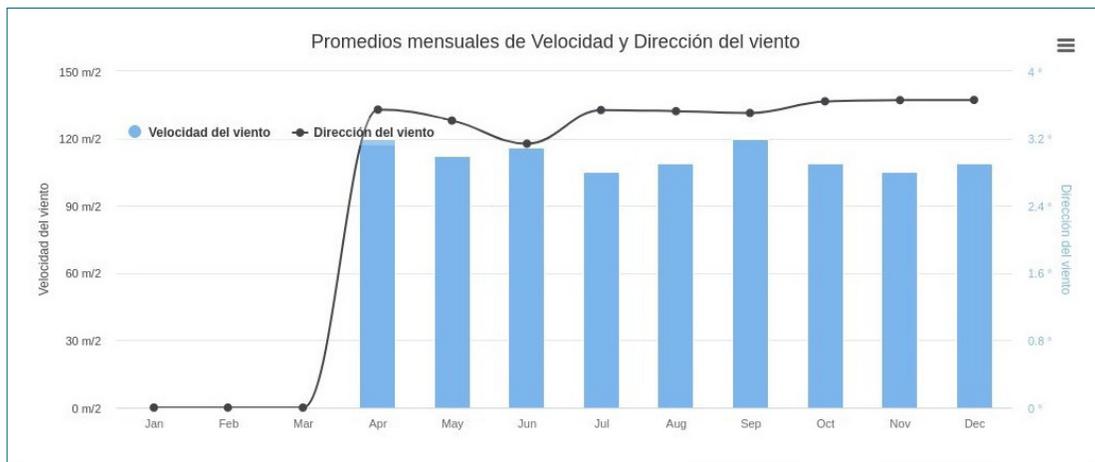
Fuente: elaboración propia.

FIGURA 2. DIAGRAMA DE LÍNEAS CON PROMEDIOS MENSUALES DE TEMPERATURA



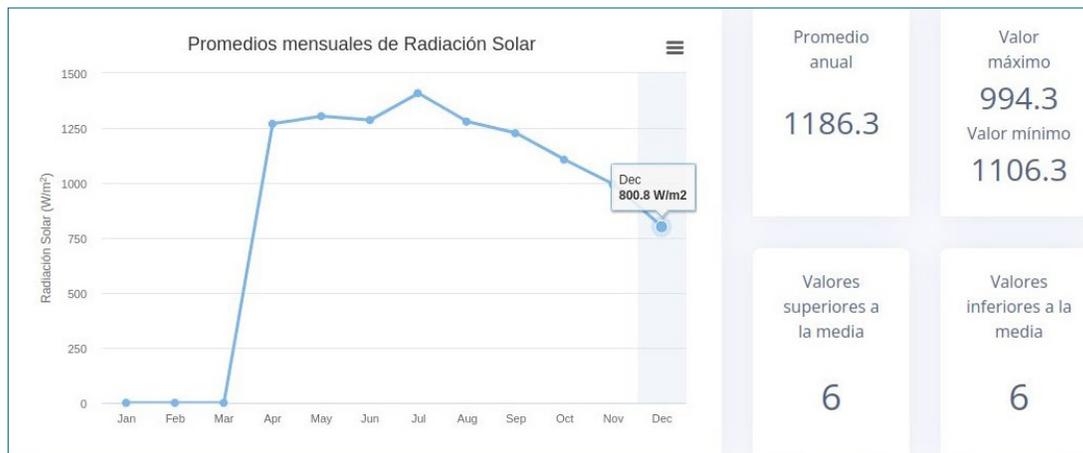
Fuente: autores.

FIGURA 3. DIAGRAMAS DE BARRAS CON PROMEDIOS MENSUALES DE LLUVIA Y HUMEDAD



Fuente: elaboración propia.

FIGURA 4. DIAGRAMA COMPUESTO CON PROMEDIOS DE VELOCIDAD Y DIRECCIÓN DEL VIENTO



Fuente: elaboración propia.

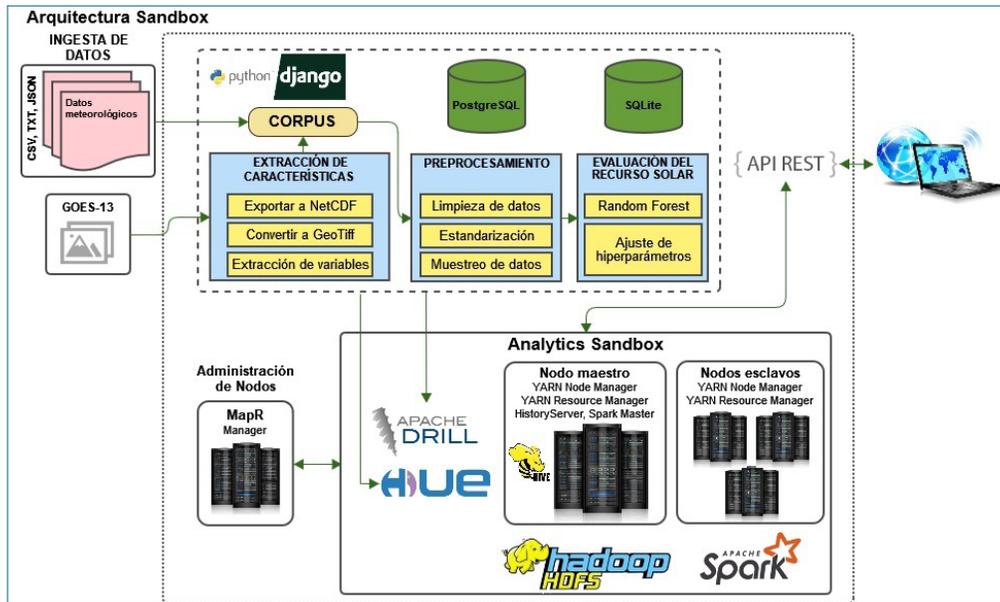
FIGURA 5. DIAGRAMA DE DISPERSIÓN CON PROMEDIOS MENSUALES DE RADIACIÓN SOLAR

El *dashboard* E-solar permite la visualización de todas las gráficas generadas históricamente. De igual manera, la librería Highcharts [26] permite la visualización de la gráfica en pantalla completa, la impresión, la descarga en diferentes formatos y la visualización de la tabla de datos.

Arquitectura general de la herramienta E-solar

El *dashboard* está implementado sobre la plataforma MapR para brindar acceso a una variedad de fuentes de datos en un clúster de computadoras, incluidas las cargas de trabajo de *big data* como Apache Hadoop [27] y Apache Spark [5], un sistema de archivos distribuido, un sistema de administración de bases de datos de múltiples modelos y procesamiento de flujo de eventos, que combina análisis en tiempo real con aplicaciones operativas.

La herramienta E-solar se alimenta de dos tipos de fuentes de información que se integran en un corpus de datos que será preprocesado y conducirá a la evaluación del recurso solar a través del algoritmo Random Forest. La figura 6 presenta el flujo de los datos del sistema sobre una arquitectura *big data* hasta la presentación de información al usuario final.



Fuente: elaboración propia.

FIGURA 6. ARQUITECTURA GENERAL DE E-SOLAR

RESULTADOS

El dashboard E-solar fue validado por un grupo de 30 profesionales de la industria energética y expertos en ciencias de la computación. Estos evaluaron su usabilidad, conformidad con los estándares de programación actuales y el perfil de usuarios que podrían beneficiarse de sus funcionalidades. La tabla 2 presenta los resultados obtenidos de cada aspecto evaluado.

TABLA 2. RESULTADOS DE LA EVALUACIÓN DE E-SOLAR

Id.	Aspecto evaluado	Sí	No
1	Necesidad de conocimientos de experto para usar el sistema.	43%	57%
2	Usuarios de diversas áreas (sistemas, eléctrica, electrónica, estadística, etc.) pueden beneficiarse del sistema.	100%	0%
3	Usuarios con diferentes niveles de experticia (novatos, intermedios, expertos) pueden aprovechar el sistema.	100%	0%
4	Accesibilidad y facilidad de uso de E-solar.	83%	17%
5	Rapidez de respuesta y uso de tecnologías adecuadas.	100%	0%

Continúa...

6	Confiabilidad y utilidad de los modelos de pronóstico o descriptivos del sistema para proyectos relacionados con el recurso solar.	83%	17%
7	Robustez del sistema por uso de algoritmos de aprendizaje automático.	100%	0%
8	Interés de otros actores fuera del sector energético.	100%	0%
9	Disposición a usar y recomendar E-solar.	100%	0%

Fuente: elaboración propia.

Los resultados de la evaluación de E-solar, sintetizados en la tabla anterior, revelan un alto nivel de aceptación y confianza entre los profesionales de la industria energética y expertos en ciencias de la computación. Destacan la versatilidad de la herramienta, con el 100 % de los participantes afirmando que usuarios de diversos campos y niveles de experticia pueden beneficiarse de ella. Aunque el 43.3 % considera necesarios conocimientos de experto, la mayoría (83.3 %) encuentra que la herramienta E-solar es accesible y fácil de usar. La confiabilidad de sus modelos predictivos (83.3 %) y la robustez otorgada por algoritmos de aprendizaje automático (100 %) son altamente valoradas. Notablemente, todos los participantes usarían y recomendarían E-solar, reconociendo su potencial más allá del sector energético.

Durante la evaluación se recibieron recomendaciones relacionadas con la implementación de funciones de asistencia dirigidas a usuarios menos experimentados, sugerencias para mejorar la interfaz gráfica del sistema y la incorporación de diferentes opciones de visualización. Además, se ha sugerido explorar la inclusión de modelos de aprendizaje profundo para mejorar la precisión de las predicciones en la evaluación del recurso solar.

En cuanto a la evaluación realizada por profesionales de destacadas empresas en la industria energética de Colombia, incluyendo a CELSIA, EMCALI, EPM, entre otras, se tuvieron en cuenta aspectos de usabilidad, su contribución a la mitigación del cambio climático y su relevancia en la utilización del recurso solar.

Las empresas del sector energético del país coinciden en que esta herramienta es valiosa para la planificación de proyectos de generación de energía basados en recursos solares. Reconocen su contribución a la mitigación del cambio climático a través de la promoción de energías renovables. Además, consideran que los datos proporcionados por la herramienta E-solar representan una alternativa confiable a los datos generados por otros modelos y a los recopilados por estaciones terrestres de monitoreo climático. También sugieren mejoras en cuanto a la visualización y el soporte para usuarios con menos experiencia. Por último, están interesados en recibir información sobre futuras actualizaciones y mejoras de la herramienta E-solar.

CONCLUSIONES

El *dashboard* E-solar está construido sobre la distribución MapR del ecosistema Hadoop. La implementación de esta arquitectura *big data* se realizó porque la herramienta utiliza datos heterogéneos como archivos planos e imágenes; además, es compatible con otras fuentes de datos como los sensores. Adicionalmente, el sistema se alimenta del uso masivo de datos, que inicialmente está en el orden de los gigabytes, sin embargo, tiene la capacidad de ser escalable a terabytes de información. Asimismo, la herramienta requiere de altas prestaciones de *hardware* para brindar redundancia en la disponibilidad de los datos y para proporcionar la velocidad suficiente en el procesamiento de la información, por lo tanto, utiliza el *framework* de código abierto PySpark [4] para la computación en paralelo sobre un clúster de 5 computadores.

Se utilizó el lenguaje de programación Python, el *framework* Django [6] y algunas librerías de JavaScript para el desarrollo de la herramienta E-solar. En Python se implementó la estructura del proyecto y los diferentes modelos de Aprendizaje Automático. El *framework* Django [6] se utilizó para el desarrollo web de la aplicación basado en el patrón de diseño modelo-vista-controlador. Finalmente, la librería Highcharts [26] de JavaScript fue utilizada para la programación de las distintas visualizaciones del *dashboard*.

El rendimiento de la aplicación cuando se hace uso de la funcionalidad de predicción de radiación solar en el territorio colombiano mejoró considerablemente. Mientras en una sola máquina, el tiempo de respuesta oscilaba en alrededor de una hora, la misma tarea ejecutada en el clúster de computadores está tardando en el orden de segundos. La funcionalidad requiere altas prestaciones de *hardware* porque necesita extraer características de alrededor de 2000 imágenes satelitales, entrenar un algoritmo y generar conjuntos de datos de radiación solar. En este sentido, el clúster de cómputo cuenta con una capacidad de almacenamiento de 4 Tb, una memoria RAM de 256 Gb y un procesamiento de datos conformado por 32 núcleos físicos y 32 núcleos lógicos.

La herramienta E-solar se constituye como una alternativa de uso para los investigadores del recurso solar y para los profesionales que implementan microrredes basadas en sistemas fotovoltaicos. Este sistema complementa la información recabada por las estaciones de observación meteorológicas del IDEAM en lo relacionado con la radiación solar. Adicionalmente, sirve como punto de comparación con los conjuntos de datos de radiación solar generados por el portal web POWER | Data Access Viewer de la NASA. También, brinda la posibilidad de realizar simulaciones para la optimización de microrredes en la herramienta Homer Pro [28].

Esta herramienta no pretende competir con el uso de datos provenientes de estaciones de observación meteorológicas del IDEAM. Tampoco existe la aspiración de reemplazar el portal web de la NASA u otros instrumentos que proveen conjuntos de datos al público en general. La intención de este sistema web es brindar una alternativa en la generación de datos para investigadores y empresarios involucrados en la evaluación del recurso solar.

Con base en los resultados de la evaluación, el *dashboard* E-solar ha recibido una evaluación positiva por parte de profesionales de la industria energética y expertos en ciencias de la computación. La herramienta se destaca por su usabilidad, conformidad con los estándares de desarrollo actuales y su capacidad para atender a usuarios de diversos niveles de experiencia. A pesar de no requerir una curva de aprendizaje pronunciada, se reconoce la utilidad de tener conocimientos básicos en análisis de datos. Las recomendaciones se centran en mejorar la asistencia a usuarios menos experimentados, la interfaz gráfica y la exploración de modelos de aprendizaje profundo.

Además, algunas de las empresas líderes en la industria energética en Colombia valoran la herramienta E-solar por su contribución a la planificación de proyectos de energía solar y la mitigación del cambio climático a través de energías renovables. Estas empresas consideran que los datos proporcionados por E-solar son confiables y representan una alternativa sólida a otras fuentes de datos. También expresan interés en mejoras futuras, particularmente en la visualización y el soporte para usuarios menos experimentados. En conjunto, los resultados reflejan una recepción positiva en la comunidad profesional y empresarial del sector energético, destacando la utilidad y relevancia de E-solar en el contexto de la generación de energía solar en Colombia.

AGRADECIMIENTOS

Deseo expresar mi agradecimiento a Minciencias, a la Universidad del Valle, a la Facultad de Ingeniería y a la Escuela de Ingeniería de Sistemas y Computación (EISC).

REFERENCIAS

- [1] X. Zheng, D. Streimikiene, T. Balezentis, A. Mardani, F. Cavallaro y H. Liao, «A review of greenhouse gas emission profiles, dynamics, and climate change mitigation efforts across the key climate change players», *Journal of Cleaner Production*, vol. 234, pp. 1113-1133, oct. 2019. doi: 10.1016/j.jclepro.2019.06.140.

- [2] L. E. Ordóñez Palacios, V. Bucheli Guerrero y H. Ordóñez, «Machine Learning for Solar Resource Assessment Using Satellite Images», *Energies*, vol. 15, n.º 11, Art. n.o 11, enero 2022. doi: 10.3390/en15113985.
- [3] NOAA, «NOAA's Office of Satellite and Product Operations», 2021. <https://www.ospo.noaa.gov/Operations/GOES/13/index.html>. [Accedido: 30 marzo 2021].
- [4] «PySpark Overview - PySpark master documentation» [En línea]. Disponible en: <https://spark.apache.org/docs/latest/api/python/index.html>. [Accedido: 7 junio 2024].
- [5] Apache Spark, «Overview - Spark 3.5.1 Documentation», 2024. <https://spark.apache.org/docs/latest/>. [Accedido: 7 junio 2024].
- [6] «Get started with Django», Django Project [En línea]. Disponible en: <https://www.djangoproject.com/>. [Accedido: 8 julio 2024].
- [7] «JavaScript | MDN» [En línea]. Disponible en: <https://developer.mozilla.org/es/docs/Web/JavaScript>. [Accedido: 8 julio 2024].
- [8] L. E. Ordóñez Palacios, «Dashboard E-Solar». 26 de noviembre de 2022. [Python]. Disponible en: <https://github.com/luise-phd/dashboard>. [Accedido: 21 marzo 2023].
- [9] NASA, «POWER | Data Access Viewer», 2021. <https://power.larc.nasa.gov/data-access-viewer/>. [Accedido: 21 marzo 2023].
- [10] Scikit-learn, «scikit-learn: machine learning in Python - scikit-learn 1.2.2 documentation», 2007. <https://scikit-learn.org/stable/>. [Accedido: 21 marzo 2023].
- [11] Developers XGBoost, «XGBoost Documentation - xgboost 1.7.4 documentation», 2022. <https://xgboost.readthedocs.io/en/stable/>. [Accedido: 22 feb. 2023].
- [12] IDEAM, «Consulta y Descarga de Datos Hidrometeorológicos», 2020. <http://dhime.ideam.gov.co/atencionciudadano/>. [Accedido: 21 marzo 2023].
- [13] C. Wu et al., «China's agricultural machinery operation big data system», *Computers and Electronics in Agriculture*, vol. 205, p. 107594, feb. 2023. doi: 10.1016/j.compag.2022.107594.
- [14] F. Rodríguez, A. Fleetwood, A. Galarza y L. Fontán, «Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control», *Renewable Energy*, vol. 126, pp. 855-864, oct. 2018. doi: 10.1016/j.renene.2018.03.070.
- [15] M. K. Boutahir, Y. Farhaoui, M. Azroul, I. Zeroual y A. El Allaoui, «Effect of Feature Selection on the Prediction of Direct Normal Irradiance | TUP Journals & Magazine | IEEE Xplore» [En línea]. Disponible en: <https://ieeexplore.ieee.org/document/9832772>. [Accedido: 7 junio 2024].
- [16] A. A. Tehrani, O. Veisi, B. V. Fakhr y D. Du, «Predicting solar radiation in the urban area: A data-driven analysis for sustainable city planning using artificial neural

- networking», *Sustainable Cities and Society*, vol. 100, p. 105042, enero 2024. doi: 10.1016/j.scs.2023.105042.
- [17] M. Cuesta, J. Constante y D. Jijón, «Modelos de Predicción de Radiación Solar y Temperatura Ambiente mediante Redes Neuronales Recurrentes», *Revista Técnica Energía*, vol. 19, pp. 81-89, enero 2023. doi: 10.37116/revistaenergia.v19.n2.2023.552.
- [18] L. Cárdenas, D. Hurtado y R. Moreno, «Predicción de radiación solar mediante deep belief network», *Revista Tecnura*, vol. 20, p. 39, enero 2016. doi: 10.14483/udistrital.jour.tecnura.2016.1.a03.
- [19] Vinod, R. Kumar y S. K. Singh, «Solar photovoltaic modeling and simulation: As a renewable energy solution», *Energy Reports*, vol. 4, pp. 701-712, nov. 2018. doi: 10.1016/j.egy.2018.09.008.
- [20] S. R. Pendem y S. Mikkili, «Modeling, simulation and performance analysis of solar PV array configurations (Series, Series-Parallel and Honey-Comb) to extract maximum power under Partial Shading Conditions», *Energy Reports*, vol. 4, pp. 274-287, nov. 2018. doi: 10.1016/j.egy.2018.03.003.
- [21] N. Abdullahi, C. Saha y R. Jinks, «Modelling and performance analysis of a silicon PV module», *Journal of Renewable and Sustainable Energy*, vol. 9, n.º 3, p. 033501, mayo 2017. doi: 10.1063/1.4982744.
- [22] Apache Hive, «LanguageManual - Apache Hive - Apache Software Foundation», 2022. <https://cwiki.apache.org/confluence/display/Hive/LanguageManual>. [Accedido: 4 abril 2023].
- [23] MapR, «HPE Ezmeral Data Fabric 7.2 Documentation», 2022. <https://docs.datafabric.hpe.com/72/install.html>. [Accedido: 20 marzo 2023].
- [24] Apache Drill, «Documentation - Apache Drill», 2022. <https://drill.apache.org/docs/>. [Accedido: 4 abril 2023].
- [25] Hue, «Quick Start : Hue SQL Assistant Documentation», 2022. <https://docs.gethue.com/quickstart/>. [Accedido: 4 abril 2023].
- [26] Highcharts, «Highcharts Documentation | Highcharts», 2020. <https://highcharts.com/docs/index>. [Accedido: 4 abril 2023].
- [27] «Apache Hadoop» [En línea]. Disponible en: <https://hadoop.apache.org/>. [Accedido: 8 julio 2024].
- [28] HOMER, «HOMER Pro - Microgrid Software for Designing Optimized Hybrid Microgrids», 2021. <https://www.homerenergy.com/products/pro/index.html> [Accedido: 9 junio 2021].