

Análisis Predictivo de la viabilidad de la instalación para plantas de Energía Solar en Colombia mediante Machine Learning y Datos Meteorológicos.

Giovanny Castillo Castillo & Andrés Felipe Sainea Rojas

Septiembre 2024.

Universidad EAN.

Departamento de Ciencia de Datos, Especialización en Machine Learning.
Seminario de Investigación.

El presente estudio tuvo como objetivo desarrollar una herramienta basada en técnicas de machine learning para evaluar la viabilidad de la instalación de plantas de energía solar en Colombia. La herramienta se diseñó para analizar datos meteorológicos históricos y en tiempo real, optimizando así la selección de ubicaciones para nuevas instalaciones solares. La metodología integró datos recolectados mediante estaciones meteorológicas y sistemas de monitoreo satelital, los cuales fueron procesados utilizando algoritmos de predicción avanzada. Este enfoque permitió mejorar la precisión de los resultados y proporcionar recomendaciones prácticas que fomentan el uso de energías renovables y aumentan la eficiencia energética en el país.

Descripción del problema

Causas u origen del problema.

En Colombia, la diversificación de fuentes energéticas es esencial debido a la alta dependencia de la energía hidroeléctrica, que representa aproximadamente el 70% de la generación eléctrica del país (UPME, 2020). Esta dependencia hace que el sistema eléctrico sea vulnerable a fenómenos climáticos extremos, como fenómeno del Niño, que disminuye significativamente los caudales de los ríos, afectando la capacidad de generación de las centrales hidroeléctricas (IDEAM, 2019). Dada esta situación, la energía solar surge como una alternativa viable para complementar la matriz energética y reducir la vulnerabilidad a la variabilidad climática (Martínez et al., 2020).

No obstante, la limitada disponibilidad de información precisa sobre las condiciones meteorológicas específicas que afectan el potencial de generación solar en distintas regiones del país ha frenado la inversión en este sector (Flórez, 2021). Pese a que Colombia cuenta con niveles elevados de radiación solar, en algunas zonas superiores a 5 kWh/m² al día (Gutiérrez, 2019), la distribución geográfica de este recurso y las condiciones climáticas variables dificultan una evaluación precisa de su potencial. Además, la infraestructura tecnológica para procesar grandes volúmenes de datos meteorológicos en el contexto de energías renovables está subdesarrollada (Ortiz, 2020).

Síntomas o situaciones anómalas.

A pesar del alto potencial de radiación solar en ciertas regiones de Colombia, que supera los 5 kWh/m² diarios (Gutiérrez, 2019), la inversión en proyectos de energía solar ha sido limitada. Esto se debe, en gran medida, a la falta de datos meteorológicos precisos y a la ausencia de modelos predictivos que evalúen de manera adecuada el potencial de generación solar a nivel regional (Flórez, 2021). Como resultado, la toma de decisiones en la instalación de plantas solares se basa en información fragmentada y estudios de viabilidad poco detallados, generando incertidumbre y riesgos elevados para los inversores (Rodríguez et al., 2022). Esto ha llevado a una subutilización del recurso solar, especialmente en regiones como la Guajira, el Huila y el Tolima, donde el crecimiento del sector solar ha sido menor al esperado (Téllez, 2021).

Pronóstico de la situación.

Si esta situación persiste, Colombia continuará dependiendo de fuentes de energía vulnerables a eventos climáticos extremos, incrementando el riesgo de inestabilidad en el suministro de energía durante periodos de sequía prolongada (Banco Mundial, 2022). La falta de aprovechamiento del potencial solar también limitará el crecimiento de las energías

renovables, dificultando la transición hacia una matriz energética más sostenible y diversificada, y manteniendo la dependencia de fuentes de energía no renovables (Pérez, 2020).

Control pronóstico

La implementación de técnicas de machine learning en el análisis de datos meteorológicos ofrece una solución para este problema. Estas técnicas permiten evaluar de manera más precisa el potencial de generación solar en diferentes regiones, considerando variables como la radiación solar, la temperatura y las condiciones climáticas locales (Ramírez, 2021). Con el desarrollo de modelos predictivos robustos, se pueden identificar las ubicaciones óptimas para la instalación de plantas solares, reduciendo la incertidumbre para los inversores y fomentando un crecimiento más acelerado del sector de energías renovables en Colombia (Flórez, 2021). Además, la capacidad de actualizar los modelos en tiempo real permite una mejor adaptación a las variaciones climáticas, optimizando así la producción energética (Ortiz, 2020).

A partir del análisis previo se plantea la siguiente pregunta de investigación:

¿Cómo puede el uso de técnicas de machine learning y el análisis de datos meteorológicos mejorar la evaluación del potencial de instalación de plantas de energía solar en Colombia, y su ubicación óptima para su implementación a gran escala?

Objetivos

Objetivo general.

Desarrollar una herramienta que permita determinar la viabilidad de la instalación de plantas de energía solar en Colombia, utilizando datos meteorológicos y técnicas de machine learning.

Objetivos específicos.

- Realizar un análisis exploratorio de los datos para identificar patrones y seleccionar las variables más relevantes que influirán en la producción de energía solar
- Desarrollar modelos de machine learning para predecir la viabilidad de instalación de plantas solares en diferentes ubicaciones
- Validar los modelos predictivos utilizando datos reales y evaluar su precisión y utilidad para la toma de decisiones estratégicas en el sector energético.

Justificación.

La importancia de realizar esta investigación radica en la necesidad de diversificación de la matriz energética de Colombia, que actualmente depende en gran

medida de la energía hidroeléctrica (UPME, 2020). La energía solar se presenta como una alternativa sostenible y viable, especialmente dadas las condiciones geográficas favorables del país, donde se registran niveles de radiación solar que superan los 5 kWh/m² diarios en varias regiones (Gutiérrez, 2019). Sin embargo, la escasez de información precisa y la ausencia de modelos predictivos basados en datos meteorológicos han limitado el aprovechamiento efectivo de este recurso (Flórez, 2021).

Por lo tanto, este trabajo busca proporcionar herramientas tecnológicas avanzadas que mejoren la toma de decisiones en la planificación e instalación de plantas solares, contribuyendo así a la resiliencia y sostenibilidad del sistema energético colombiano.

Este estudio se alinea con el campo de Ciencia, Tecnología e Innovación, y se enmarca dentro del grupo INDEVOS de la Universidad Ean, específicamente en la línea de investigación de Soluciones basadas en la naturaleza y cambios transformativos hacia la sostenibilidad, enfocándose en Energías Renovables y Eficiencia Energética (EREE).

Este estudio es conveniente porque responde a las metas de sostenibilidad energética trazadas tanto por Colombia como por la comunidad internacional en los Objetivos de Desarrollo Sostenible (ONU, 2015). La capacidad de aprovechar el potencial solar a través de técnicas de machine learning contribuirá a la descarbonización de la matriz

energética, reduciendo la dependencia de fuentes no renovables y los efectos negativos del cambio climático (Pérez, 2020).

El acceso a fuentes de energía limpias y sostenibles tiene un impacto directo en la calidad de vida de la población colombiana. Una mejor integración de la energía solar en la matriz energética contribuirá a reducir la frecuencia de apagones y la inestabilidad del suministro energético, especialmente en épocas de sequía. Además, el desarrollo de la energía solar en regiones con alto potencial, como la Guajira y el Huila, podría fomentar el crecimiento económico local, creando empleos y promoviendo el desarrollo en comunidades rurales y marginadas (Téllez, 2021). De esta manera, la investigación no solo promueve la sostenibilidad ambiental, sino también la equidad social y el bienestar comunitario.

Este estudio propone el uso de modelos predictivos basados en machine learning para evaluar de manera precisa el potencial solar de diferentes regiones del país. Esto permitirá a los tomadores de decisiones en el sector energético contar con información confiable y actualizada para planificar la instalación de plantas solares. La disponibilidad de datos precisos reduce la incertidumbre para los inversores, facilitando la asignación de recursos y la identificación de las mejores ubicaciones para proyectos solares. Como resultado, se espera un incremento en la inversión en energía solar, que fortalecerá la

resiliencia del sistema eléctrico colombiano y acelerará la adopción de tecnologías limpias (Rodríguez et al., 2022).

Desde un enfoque teórico, esta investigación aporta al campo de la inteligencia artificial aplicada a las energías renovables, específicamente en contextos de países en desarrollo. El uso de algoritmos de machine learning para el análisis de datos meteorológicos ofrece una forma innovadora de evaluar el potencial solar, lo que puede ser replicable en otras regiones con características similares. Al validar y perfeccionar estos modelos en un contexto único como el colombiano, la investigación no solo amplía el conocimiento sobre la integración de la energía solar en matrices energéticas, sino que también contribuye a la literatura sobre el uso de machine learning en el sector energético (Ramírez, 2021).

Metodológicamente, este estudio ofrece un enfoque innovador y fácilmente replicable para el análisis de grandes volúmenes de datos meteorológicos en tiempo real. La metodología propuesta combina técnicas avanzadas de machine learning y Big Data, lo que permite procesar y analizar de manera eficiente datos climáticos complejos. Esto no solo incrementa la precisión de las predicciones sobre el potencial de generación de energía solar, sino que también facilita una toma de decisiones más ágil y fundamentada en el sector energético (Ortiz, 2020).

El enfoque metodológico empleado en este estudio tiene el potencial de ser adaptado y aplicado en otras regiones del mundo con características similares, contribuyendo al desarrollo de proyectos de energía solar en distintos contextos. Esto lo convierte en una herramienta valiosa para investigadores y profesionales que buscan optimizar la integración de energías renovables en la matriz energética de sus respectivos países.

Marco teórico

Estado del Arte

Energía Solar

La energía solar es el proceso de transformación de la radiación solar en energía eléctrica mediante el uso de paneles fotovoltaicos. Colombia tiene un potencial

significativo para la generación de energía solar porque tiene áreas que reciben más de 5 kWh/m² diarios de radiación solar, más que el promedio mundial. Martínez et al. (2020) destacan esto.

La energía solar ha captado la atención de la comunidad científica y el sector energético a nivel global, siendo vista como una solución viable para reducir la dependencia de los combustibles fósiles y mitigar el cambio climático. El suroeste de Estados Unidos y el norte de África han sido ejemplos paradigmáticos del potencial de la energía solar, gracias a sus altos niveles de radiación y políticas gubernamentales que han incentivado su desarrollo (Martínez et al., 2020; Rodríguez et al., 2022). En América Latina, Chile y Brasil han emergido como líderes en la implementación de proyectos solares a gran escala, impulsados por marcos regulatorios favorables y la disminución de costos tecnológicos en la producción de paneles fotovoltaicos.

Colombia, bendecida con una ubicación geográfica que la convierte en un paraíso para la producción de energía solar, ha avanzado de manera más discreta en la implementación de proyectos de gran envergadura. Aunque estudios como los de Flórez (2021) y Pérez (2020) revelaron los altos niveles de radiación solar en el país, aún no se ha profundizado en un análisis exhaustivo que considere las variables meteorológicas específicas de cada región. En otros países, como Alemania y Australia, la inteligencia artificial ha emergido como un faro de innovación, utilizando el poder del machine learning

para procesar vastos volúmenes de datos meteorológicos y optimizar la integración de la energía solar en sus sistemas energéticos, según Ahmed et al. (2020) y Li et al. (2019). En este contexto, la aplicación de modelos basados en machine learning ha demostrado ser una herramienta invaluable para la gestión energética y la predicción precisa de la generación de energía solar, incluso en microrredes conectadas a la red eléctrica que integran diversas fuentes de energía distribuida, como lo expresa el artículo "Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources" (Ahmed et al., 2020).

A nivel internacional, países como Alemania y Australia han utilizado con éxito algoritmos de inteligencia artificial para el análisis de datos meteorológicos, permitiendo una mayor precisión en la predicción del rendimiento de las plantas solares (Ahmed et al., 2020; Li et al., 2019). Este enfoque no solo optimiza la producción energética, sino que también facilita la integración de la energía solar en las redes eléctricas nacionales, mitigando las incertidumbres asociadas con la variabilidad climática. Para Colombia, donde la intermitencia de la energía hidroeléctrica plantea riesgos debido a los fenómenos meteorológicos extremos, la diversificación de la matriz energética mediante fuentes renovables como la solar es esencial (Cárdenas y Pérez, 2020).

Sostenibilidad Energética

La capacidad de un sistema para satisfacer las necesidades energéticas actuales sin comprometer las futuras generaciones se conoce como sostenibilidad energética. Según Sovacool (2016), la transición hacia una matriz energética más sostenible, en la que la energía solar es crucial, es esencial para mitigar los efectos negativos del cambio climático y garantizar la seguridad energética a largo plazo.

Teorías y modelos conceptuales sobre la energía y modelos de Machine learning.

Teoría de la Transición Energética

La teoría de la transición energética, propuesta por Sovacool (2016), ofrece un marco conceptual para entender cómo las sociedades pueden migrar de sistemas energéticos basados en combustibles fósiles hacia sistemas más sostenibles como los basados en energía solar, eólica y geotérmica. En el contexto colombiano, esta transición no solo es necesaria para enfrentar el cambio climático, sino que también es vital para garantizar la seguridad energética en el largo plazo. El país, cuya matriz energética depende en gran medida de la energía hidroeléctrica, enfrenta serios riesgos ante eventos como el

fenómeno de El Niño, que puede reducir drásticamente la capacidad de las represas. La inclusión de la energía solar en el mix energético colombiano puede mejorar la resiliencia del sistema y reducir la vulnerabilidad ante estas fluctuaciones climáticas (Cárdenas y Pérez, 2020).

El éxito de esta transición no solo depende de la disponibilidad de recursos naturales como la radiación solar, sino también de un entorno favorable en términos de políticas públicas, infraestructura tecnológica y la aceptación social. En este sentido, la teoría de Sovacool subraya la importancia de una planificación a largo plazo que contemple no solo las variables económicas y tecnológicas, sino también el impacto social y ambiental que este cambio generará (Sovacool, B. 2016).

Modelos Predictivos en Energía Renovable.

La predicción del potencial de energía solar mediante modelos predictivos ha sido un campo que ha florecido en los últimos años. Ramírez (2021) destaca la capacidad de estos modelos para generar pronósticos precisos, utilizando variables climáticas como la radiación solar y la temperatura. En países como Alemania, estos modelos han permitido integrar de manera eficiente la energía solar en las redes eléctricas, optimizando tanto la producción como la distribución de energía (Ahmed et al., 2020). Este proceso no solo ha

significado un avance técnico, sino también un triunfo del ingenio humano, que busca armonizarse con la naturaleza para forjar un futuro más limpio y próspero.

Machine Learning

En el campo de la energía solar, la aplicación de algoritmos permite a los sistemas aprender y mejorar sus predicciones a partir de grandes conjuntos de datos. Según Ahmed et al. (2020), las redes neuronales y las SVM son ejemplos de técnicas de aprendizaje automático que han optimizado significativamente la predicción de la radiación solar. Esto mejora la toma de decisiones estratégicas sobre la ubicación y tamaño de las plantas solares.

Modelos de Machine Learning más utilizados.

Redes Neuronales Artificiales (ANNs)

El poder de las redes neuronales artificiales radica en su capacidad para descubrir patrones no lineales en grandes conjuntos de datos, algo que ha transformado por completo la predicción del potencial energético. Li et al. (2019) destacan que estas redes han sido utilizadas en todo el mundo para prever con precisión la generación de energía y la radiación solar, adaptando sus predicciones a medida que nuevos datos enriquecen los modelos. Este enfoque flexible, cargado de posibilidades, subraya el impacto transformador que la inteligencia artificial está teniendo en la sostenibilidad energética.

Máquinas de Soporte Vectorial (SVMs)

Las máquinas de soporte vectorial han demostrado ser igualmente valiosas en la predicción de la radiación solar. Ahmed et al. (2020) muestran cómo estos modelos no solo procesan variables climáticas con una precisión sorprendente, sino que también son capaces de manejar datos complejos de alta dimensionalidad. Estas herramientas se han convertido en faros de esperanza para quienes buscan soluciones más precisas y adaptadas a las realidades de cada región. Además, se ha demostrado que el uso de diversos modelos de machine learning, ajustados mediante algoritmos avanzados como el *chimp optimization algorithm*, optimizan significativamente la predicción de la producción de energía fotovoltaica.

Bosques Aleatorios (Random Forests)

Los bosques aleatorios son modelos de aprendizaje por ensamble que combinan varios árboles de decisión para mejorar la precisión predictiva. Estos modelos son especialmente útiles en el análisis de datos meteorológicos, ya que pueden manejar grandes cantidades de datos sin caer en el sobreajuste (Flórez, 2021). Su capacidad para procesar datos de diferentes fuentes y evitar errores comunes en el modelado tradicional los convierte en una de las mejores herramientas para predecir la viabilidad del uso de la energía solar en diversos climas.

Redes Neuronales Recurrentes (RNNs)

Las redes neuronales recurrentes (RNN) son fundamentales en la predicción de series temporales, como las fluctuaciones diarias y estacionales de la radiación solar. Martínez et al. (2020) señalan que las RNN, particularmente aquellas con LSTM (*Long short-term memory*), han mostrado ser altamente efectivas en la predicción de la generación de energía solar en entornos donde las condiciones climáticas varían significativamente.

Regresión Lineal y Polinomial

Aunque las técnicas de regresión lineal y polinomial han sido tradicionalmente utilizadas para la predicción de la radiación solar, su simplicidad las hace menos precisas en comparación con los modelos más avanzados de machine learning. Sin embargo, siguen siendo una herramienta útil en las primeras etapas de planificación, cuando se requiere obtener estimaciones rápidas y relativamente precisas sin la necesidad de realizar análisis complejos (Gutiérrez, 2019).

Máquinas de aprendizaje profundo (*Deep Learning*)

Las máquinas de aprendizaje profundo han ganado notoriedad en la predicción de energía solar gracias a su capacidad para procesar y analizar grandes volúmenes de datos. Este enfoque utiliza arquitecturas complejas, como las redes neuronales convolucionales (CNN), que pueden identificar características significativas en los datos meteorológicos y de irradiación solar. Bhandari y Vardhan (2021) enfatizan que el aprendizaje profundo

permite una mayor precisión en la predicción de la generación de energía solar, incluso en condiciones altamente variables, gracias a su capacidad para captar patrones complejos en los datos. Este enfoque se ha convertido en un recurso valioso para la optimización de proyectos de energía solar en diversas regiones.

Marco Legal y Regulatorio en Energía Solar en Colombia

La Ley 1715 de 2014 es un pilar fundamental en el marco legal que busca promover la integración de las energías renovables en el sistema eléctrico colombiano. Esta ley establece las bases para la promoción y desarrollo de fuentes de energía no convencionales, incluyendo la energía solar, y su objetivo principal es diversificar la matriz energética del país, que históricamente ha estado dominada por la energía hidroeléctrica.

La ley fomenta la inversión en proyectos de energía renovable mediante incentivos económicos, como beneficios tributarios y mecanismos de financiación. Sin embargo, a pesar de los avances significativos que esta legislación ha traído, la implementación de proyectos solares a gran escala aún enfrenta varios desafíos.

Uno de los principales obstáculos en la expansión de la energía solar en Colombia es la falta de incentivos sólidos y claros para los inversionistas. Aunque se han implementado políticas para fomentar la inversión, como la Ley 1715 de 2014 y su

modificación con la Ley 2099 de 2021, la incertidumbre regulatoria persiste. Según Flórez (2021), los procedimientos regulatorios son complejos, lo que ha generado confusión y desmotivación entre los posibles inversionistas. Además, a pesar de que la Comisión de Regulación de Energía y Gas (CREG) ha establecido mecanismos de subasta para la compra de energía renovable, la falta de claridad en los plazos y condiciones ha limitado su aprovechamiento (ANDI, 2023).

Otro desafío importante es la modernización de la infraestructura tecnológica del país. Las redes eléctricas actuales requieren mejoras para manejar la intermitencia de la energía solar y facilitar la integración de nuevos proyectos. La implementación de tecnologías avanzadas, como los sistemas de almacenamiento de energía, es esencial para asegurar un suministro energético confiable (ANDI, 2023). Por otro lado, aunque se han creado incentivos como el Registro de Generación Distribuida, su implementación efectiva sigue siendo un reto importante (Estudio Legal Hernández, 2023)

Además de la Ley 1715, otras normativas y políticas complementarias han surgido para respaldar el desarrollo de energías renovables. El Plan Nacional de Desarrollo incluye objetivos claros en cuanto a la promoción de energías limpias y sostenibles. También se han creado incentivos regionales en algunos departamentos para facilitar la instalación de sistemas de energía solar, destacando el potencial de regiones como La Guajira, que posee condiciones ideales para la generación solar.

A medida que Colombia avanza hacia una matriz energética más diversificada y sostenible, es imperativo que el marco legal y regulatorio evolucione para abordar los desafíos actuales y fomentar un entorno favorable para la inversión en energía solar. Esto no solo beneficiará la economía del país, sino que también contribuirá a la lucha contra el cambio climático y a la seguridad energética a largo plazo.

Metodología

Enfoque de Investigación

El estudio utiliza un enfoque mixto, combinando métodos cuantitativos y cualitativos para evaluar el potencial de generación de energía solar en diversas regiones de Colombia. Desde la perspectiva cuantitativa, se basa en la recolección y análisis de datos meteorológicos cuantificables, permitiendo identificar patrones y realizar predicciones sobre la viabilidad energética. En cuanto al análisis cualitativo, este se incorpora después de las técnicas de análisis de información, con el objetivo de interpretar los resultados obtenidos y contextualizarlos dentro de las condiciones específicas de cada región. El diseño de la investigación es de tipo no experimental, ya que no se manipulan variables, y de carácter transversal, dado que los datos se recolectaron en un único punto en el tiempo. Este enfoque mixto permite desarrollar modelos predictivos robustos que no solo determinen la viabilidad de instalar plantas de energía solar, sino que también ofrezcan una interpretación más profunda de los resultados en función de las características socioambientales de las regiones analizadas.

Diseño de la Investigación y Alcance

Este proyecto se enmarca en un estudio correlacional y descriptivo, ya que busca identificar y describir las relaciones entre variables meteorológicas como la radiación solar, la temperatura y la humedad relativa, y su impacto en la generación de energía solar. A través

de la aplicación de técnicas de machine learning, se pretende analizar grandes conjuntos de datos históricos y en tiempo real, con el fin de optimizar la toma de decisiones estratégicas para la instalación de plantas solares.

Definición de Variables

- **Radiación Solar:** Se define como la cantidad de energía solar recibida en una superficie específica, medida en kWh/m² diarios. Se considera una de las variables clave para determinar el potencial de generación de energía solar.
- **Temperatura Ambiental:** La temperatura promedio de cada región es un factor que influye en la eficiencia de los paneles solares.
- **Humedad Relativa:** Esta variable se mide para entender su posible impacto en la variabilidad de la radiación solar.

Cada variable será medida utilizando datos proporcionados por el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia) y otras fuentes confiables de datos meteorológicos.

Análisis de datos meteorológicos históricos para identificar el potencial de generación de energía solar

Para la investigación se emplean los registros meteorológicos de las regiones de Colombia con mayor potencial solar, incluyendo la Guajira, el Huila, y el Tolima. Se utilizará un muestreo de conveniencia, seleccionando aquellos registros que abarquen un periodo significativo de tiempo (mínimo 10 años de datos) para garantizar la robustez de los modelos predictivos.

Instrumentos y Recolección de Datos

En esta investigación descriptiva con enfoque correlacional, se utilizaron datos meteorológicos y geográficos previamente validados y recopilados por organizaciones internacionales y nacionales. La validación de los datos se llevó a cabo mediante procedimientos que garantizan su calidad, consistencia y relevancia para el análisis. Esto incluyó la comparación de registros históricos entre diferentes fuentes, como el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), la NASA y bases de datos públicas de energía solar, asegurando que los valores fueran consistentes y libres de discrepancias significativas. Además, se verificó la precisión de los datos mediante métodos estadísticos, como la detección y eliminación de valores atípicos, y se corroboró su correspondencia con estándares internacionales en la medición de variables climáticas. Estos datos proporcionaron la información necesaria para medir las variables clave como

la radiación solar, la temperatura y otros factores climáticos que influyen en la viabilidad de la instalación de plantas de energía solar en Colombia.

Debido a que este estudio no incluye directamente la recolección de nuevos datos mediante entrevistas o encuestas, no será necesario diseñar nuevos instrumentos cualitativos como formatos de observación o entrevistas estructuradas. Sin embargo, si fuera necesario en fases posteriores de la investigación, se diseñarán dichos instrumentos siguiendo criterios estandarizados de validez y confiabilidad, asegurando su coherencia con las variables a medir y los objetivos planteados.

Para el análisis de los datos meteorológicos y geográficos obtenidos, se emplearán técnicas de *big data* y *machine learning*. Los datos serán preprocesados, limpiados y normalizados para asegurar su calidad antes de ser integrados en los modelos predictivos. Las herramientas tecnológicas empleadas incluyen Python y bibliotecas especializadas, como Scikit-Learn y TensorFlow, las cuales son ampliamente utilizadas para el manejo y análisis de grandes volúmenes de datos.

Entre los algoritmos de *machine learning* utilizados, se destacan:

- **Redes Neuronales Artificiales (ANN):** Estas redes son capaces de detectar patrones complejos en los datos meteorológicos y geográficos, lo que las convierte en herramientas clave para la predicción en sistemas no lineales (Li et al., 2019).

- **Bosques Aleatorios (Random Forest):** Este algoritmo es adecuado para trabajar con datos no lineales y generar predicciones precisas al combinar múltiples árboles de decisión, lo que mejora su robustez frente a datos variados (Ahmed et al., 2020).
- **Máquinas de Soporte Vectorial (SVM):** Las SVM son especializadas en la clasificación de grandes conjuntos de datos, proporcionando alta precisión en la predicción de resultados, especialmente en problemas con múltiples dimensiones (Bhandari & Vardhan, 2021).

Estos modelos se seleccionaron debido a su robustez en la detección de patrones y tendencias que permitan desarrollar una herramienta predictiva eficaz para evaluar la viabilidad de instalar plantas solares en diversas regiones de Colombia.

Desarrollo de modelos de machine learning para predecir la viabilidad de instalación de plantas solares en diferentes ubicaciones

Una vez obtenidos los datos meteorológicos y geográficos a través de fuentes secundarias, como bases de datos públicas y agencias gubernamentales, se aplicarán diferentes técnicas para su análisis con el fin de generar resultados y responder a los objetivos de la investigación. A continuación, se mencionan los instrumentos de recolección, las técnicas de análisis utilizadas y una breve descripción de su aplicación:

Tabla 1.**Instrumentos de recolección de datos y técnicas de análisis empleadas.**

Instrumento	Técnica de análisis	Descripción
Datos meteorológicos (IDEAM, NASA)	Estadística descriptiva	Resumen de las variables clave como radiación solar, temperatura y humedad relativa (medias, medianas, etc.)
Datos geográficos (coordenadas, altitud).	Análisis de regresión lineal múltiple.	Identificación de las variables clave como radiación solar y los niveles de eficiencia energética solar.
Modelos de machine learning (Python).	Redes Neuronales Artificiales (ANN), Random Forest, SVM.	Predicción de la viabilidad de las plantas solares en función de patrones no lineales detectados en los datos.

Datos históricos de producción solar.	Análisis de series temporales.	Evaluación de variabilidad en la producción de tendencias futuras.
---------------------------------------	--------------------------------	--

Tabla 1, Fuente, elaboración propia.

En la investigación sobre el desarrollo de una herramienta para validar la instalación de plantas de energía solar en Colombia, es fundamental aplicar técnicas de análisis de datos que permitan interpretar y extraer información significativa de los datos recolectados. Estas técnicas no solo facilitan la identificación de patrones y relaciones entre las variables, sino que también ayudan a validar las hipótesis planteadas en el estudio. A continuación, se describen con mayor detalle las técnicas de análisis que se emplearán.

1. Estadística descriptiva

La estadística descriptiva es el primer paso en el análisis de datos, proporcionando un resumen claro y comprensible de las características básicas de los datos recolectados. Esta técnica se centra en describir y presentar los datos de manera que se pueda identificar fácilmente la tendencia central, la dispersión y la forma de la distribución de las variables (Montgomery et al., 2020).

Aplicación

En el contexto de la energía solar, la estadística descriptiva se utilizará para analizar datos meteorológicos como la radiación solar, temperatura y precipitación, y permitirá calcular:

Medidas de tendencia central: Estas incluyen la media, mediana y moda, que ayudarán a entender el comportamiento general de las variables meteorológicas y su influencia en la producción solar. Por ejemplo, la media de la radiación solar diaria en diferentes regiones puede indicar cuál es la zona más favorable para la instalación de paneles solares (Rodríguez et al., 2021).

Medidas de dispersión: La desviación estándar y el rango intercuartílico se usarán para evaluar la variabilidad de las mediciones de radiación solar y temperatura. Esto es crucial, ya que una alta variabilidad puede implicar que no todas las instalaciones serán igualmente efectivas (Smith, 2022). También se calcularán percentiles para entender mejor cómo se distribuyen los datos y si existen valores atípicos que podrían influir en los análisis posteriores.

Distribuciones de frecuencia: Estas permitirán visualizar la frecuencia de diferentes niveles de radiación solar, lo cual es útil para la toma de decisiones sobre la ubicación de nuevas plantas solares. A través de histogramas y gráficos de barras, se podrán

identificar las horas del día o las estaciones del año con mayor irradiación (Gómez et al., 2023).

Herramientas

Se emplearán herramientas como Excel y software estadístico como SPSS o R para realizar estos análisis. Además, se generarán gráficos y tablas que faciliten la visualización de los datos, permitiendo a los tomadores de decisiones comprender rápidamente la situación (López et al., 2021). Estas visualizaciones ayudarán a resumir la información de manera que sea accesible para diversas partes interesadas, incluidas autoridades gubernamentales y empresas privadas interesadas en inversiones en energía solar.

2. Análisis de regresión lineal múltiple

La regresión lineal múltiple es una técnica estadística avanzada que permite examinar la relación entre una variable dependiente y múltiples variables independientes. En este estudio, la variable dependiente será la eficiencia de producción de energía solar, mientras que las variables independientes incluirán factores geográficos y meteorológicos.

Aplicación

Esta técnica se utilizará para determinar cómo las variables como la altitud, la inclinación del terreno y la radiación solar interactúan entre sí y afectan la producción de energía solar. Por ejemplo, el análisis puede revelar que, aunque la radiación solar es un

factor crítico, la altitud también juega un papel importante en la eficiencia de los paneles solares debido a la reducción de la atmósfera (Jiménez, 2020).

La regresión permitirá:

Identificar factores significativos: A través de pruebas de hipótesis, se podrá determinar cuáles variables tienen un impacto significativo en la producción energética, lo que guiará las recomendaciones para futuras instalaciones (Martínez & Ruiz, 2021). Se empleará un nivel de significancia ($p < 0.05$) para establecer la relevancia estadística de las variables.

Predecir la producción energética: A partir del modelo de regresión, se podrán realizar predicciones sobre la producción energética en diferentes condiciones. Este aspecto es crucial para la planificación de la capacidad instalada y para evaluar si la inversión en nuevas plantas es viable (Torres & Suárez, 2023).

Evaluar la multicolinealidad: Se utilizarán índices como el Factor de Inflación de Varianza (VIF) para comprobar la multicolinealidad entre las variables independientes, asegurando que cada variable contribuya de manera única al modelo (Pérez, 2020).

Herramientas

El análisis se llevará a cabo utilizando software estadístico como R o Python, que permiten la ejecución de modelos de regresión complejos y la validación de supuestos como la normalidad, homocedasticidad y multicolinealidad (Pérez, 2020). Esto asegurará que los resultados obtenidos sean robustos y fiables.

3. Algoritmos de Machine Learning

El uso de algoritmos de machine learning proporciona un enfoque innovador para el análisis de datos en este estudio. Estas técnicas son especialmente útiles cuando se trabaja con grandes volúmenes de datos y se busca identificar patrones complejos que no son evidentes a través de métodos estadísticos tradicionales.

Aplicación

Los algoritmos seleccionados para este estudio incluyen:

Redes Neuronales Artificiales (ANN): Las ANN son útiles para modelar relaciones no lineales y pueden manejar interacciones complejas entre múltiples variables. Se entrenarán con datos históricos de producción solar y variables meteorológicas para predecir la eficiencia de nuevas instalaciones solares (Rojas & Fernández, 2019). Este método es especialmente potente cuando se dispone de grandes cantidades de datos, lo que permite que la red "aprenda" patrones en los datos de entrenamiento y realice predicciones precisas en datos no vistos.

Bosques Aleatorios (Random Forest): Este algoritmo proporciona una forma robusta de abordar problemas de clasificación y regresión. Se utilizará para identificar las características más importantes que influyen en la producción solar y para realizar predicciones basadas en conjuntos de datos más amplios y diversos (García et al., 2022). La capacidad de este algoritmo para manejar datos faltantes y su resistencia al sobreajuste lo convierten en una opción favorable en este contexto.

Máquinas de Soporte Vectorial (SVM): Las SVM se emplearán para clasificar áreas en términos de viabilidad para la instalación de plantas solares. Esta técnica es eficiente en espacios de alta dimensión y puede encontrar un hiperplano óptimo que separe las clases (López, 2021). Esto permitirá identificar regiones que cumplen con los criterios técnicos necesarios para la instalación de paneles solares, considerando factores como la radiación solar y el terreno.

Herramientas

Se utilizará Python como lenguaje principal, aprovechando bibliotecas como Scikit-Learn y TensorFlow para implementar estos algoritmos. La preparación de datos incluirá la limpieza y normalización de estos para asegurar que los modelos se ajusten correctamente (Rodríguez et al., 2021). Además, se implementarán técnicas de validación cruzada para garantizar que los modelos no sobreajusten a los datos de entrenamiento y generalicen bien a nuevos datos.

4. Análisis de series temporales

El análisis de series temporales es una técnica estadística que se utiliza para analizar datos recolectados a lo largo del tiempo, lo que permite identificar tendencias y patrones estacionales. En este estudio, esta técnica se aplicará a los datos históricos de producción solar.

Aplicación

El análisis de series temporales ayudará a:

Identificar tendencias a largo plazo: Se buscarán patrones en los datos históricos que indiquen si la producción solar está aumentando o disminuyendo con el tiempo. Esto es crucial para prever la viabilidad de nuevas instalaciones en función de cambios estacionales o climáticos (López et al., 2021). La identificación de tendencias también puede ser útil para la planificación estratégica, permitiendo a las empresas anticipar la demanda futura de energía.

Realizar pronósticos: Utilizando modelos como ARIMA (Autoregressive Integrated Moving Average), se podrán realizar pronósticos sobre la producción futura de energía solar en diferentes regiones de Colombia. Este tipo de análisis es esencial para la planificación de la capacidad instalada y para la optimización de la producción energética

(Torres & Suárez, 2023). Los pronósticos permitirán ajustar la operación de las plantas y la distribución de recursos.

Detectar patrones estacionales: Se analizarán las variaciones estacionales que pueden influir en la producción de energía solar, como la diferencia en la radiación solar entre estaciones (Gómez et al., 2023). Esta información es valiosa para la gestión de la producción energética, permitiendo ajustar la operación de las plantas solares en función de la demanda anticipada.

Herramientas

Se utilizarán software estadísticos y bibliotecas de Python, como Statsmodels, para ejecutar el análisis de series temporales, que permitirá descomponer las series en componentes estacionales, de tendencia y de ruido, lo que facilitará la interpretación de los resultados (Gómez et al., 2023). Además, se generarán gráficos de series temporales que permitirán visualizar claramente las fluctuaciones a lo largo del tiempo.

5. Análisis cualitativo.

Aunque el enfoque principal de este estudio es cuantitativo, se propone como una posible línea de investigación futura un análisis cualitativo que complemente los hallazgos obtenidos mediante los métodos cuantitativos. Este análisis cualitativo podría proporcionar

una visión más rica y contextualizada del panorama de la energía solar en Colombia, abriendo nuevas perspectivas para futuras investigaciones.

En esta etapa futura, la recolección de datos cualitativos se llevaría a cabo a través de entrevistas semiestructuradas y grupos focales con actores clave en el sector de la energía solar, como inversionistas, expertos en energía renovable y representantes gubernamentales. Estas técnicas permitirían explorar en profundidad las percepciones, actitudes y experiencias de los participantes respecto a la instalación de plantas solares en el país.

Las entrevistas semiestructuradas se realizarían a un grupo de aproximadamente 15 a 20 participantes, seleccionados por su experiencia y relevancia en el sector energético, ofreciendo flexibilidad para profundizar en temas específicos que surjan durante la conversación. Esto podría arrojar información valiosa sobre las barreras y oportunidades para la inversión en energía solar en Colombia (Hernández & Pérez, 2022). Por otro lado, los grupos focales reunirían a 3 grupos de 6 a 8 participantes cada uno, fomentando la interacción entre los participantes, lo que permitiría que se generaran dinámicas grupales que enriquecerían el análisis (González et al., 2021).

Además, el análisis del discurso se consideraría como una técnica adicional para esta fase futura, con el objetivo de explorar cómo los actores clave comunican sus

perspectivas sobre la energía solar y cómo estas narrativas afectan la percepción pública y las decisiones gubernamentales (López & Martínez, 2020). Este enfoque cualitativo representaría un "next step" en la profundización del estudio, fortaleciendo las conclusiones y proponiendo recomendaciones más integrales.

Procedimientos

- Recolección de datos meteorológicos históricos.
- Limpieza y preprocesamiento de los datos para eliminar valores atípicos y normalizar las variables.
- Aplicación de modelos de machine learning para predecir la generación de energía solar en diferentes escenarios climáticos.
- Validación de los modelos con datos de años recientes para asegurar su robustez.
- Generación de reportes y mapas de las regiones con mayor viabilidad para la instalación de plantas solares.

Validación de los Modelos Predictivos Utilizando Datos Reales

En esta etapa, los modelos predictivos desarrollados se validarán mediante la aplicación de datos reales provenientes de las fuentes previamente mencionadas (como el IDEAM, la NASA, y bases de datos de energía solar). El proceso de validación incluirá los siguientes pasos:

Selección de Datos Reales: Se seleccionarán conjuntos de datos representativos de las variables meteorológicas y geográficas relevantes para la generación de energía solar, recolectados durante un periodo de tiempo determinado. Los datos utilizados en la validación serán consistentes con los datos empleados en el entrenamiento de los modelos.

Aplicación de Modelos a Datos Reales: Los modelos predictivos serán aplicados a los datos reales para generar predicciones de producción de energía solar. Las predicciones se compararán con los valores observados de generación de energía en las zonas seleccionadas.

Evaluación de la Precisión: Se evaluará la precisión de las predicciones utilizando métricas estadísticas estándar, como el error cuadrático medio (RMSE), el error absoluto medio (MAE), y el coeficiente de determinación R^2 . Estas métricas permiten medir la diferencia entre los valores predichos y los valores reales, proporcionando una visión clara de la exactitud del modelo.

Ajuste de Modelos: En caso de que se identifiquen discrepancias significativas entre las predicciones y los datos reales, los modelos serán ajustados utilizando técnicas como la recalibración de parámetros y la inclusión de nuevas variables que puedan mejorar la precisión y relevancia de las predicciones.

Informe de Resultados: Finalmente, se documentarán los resultados de la validación en un informe detallado que incluirá una comparación entre los modelos predictivos y los datos reales, una evaluación de su precisión, y un análisis de su aplicabilidad para decisiones estratégicas en el sector energético.

Ética y Consideraciones Legales

El estudio respetará todas las normativas relacionadas con el uso de datos meteorológicos y la privacidad de la información. Asimismo, se seguirán las directrices de la Universidad EAN para el manejo de información sensible y el uso ético de herramientas tecnológicas en la investigación.

Análisis y discusión de los resultados

Introducción

El presente análisis se centra en evaluar los resultados obtenidos mediante la aplicación de un modelo de clasificación basado en Random Forest, diseñado para clasificar la viabilidad de instalación de plantas solares en Colombia. Las variables consideradas en el estudio incluyen irradiancia solar, humedad relativa, temperatura ambiental, latitud y longitud, las cuales se analizaron para determinar su influencia en la viabilidad energética. Este modelo predice tres niveles de viabilidad (Baja, Media, Alta), proporcionando información crucial para la planificación estratégica en el sector energético.

En esta sección, se discuten los resultados obtenidos, su interpretación a la luz de los objetivos del estudio y su alineación con la literatura académica. Los resultados se presentan de manera organizada según los objetivos específicos, considerando las variables estudiadas y sus dimensiones.

Clasificación de Viabilidad Energética

El modelo Random Forest mostró un desempeño adecuado en la clasificación de las regiones de Colombia según su nivel de viabilidad energética para la instalación de plantas solares. Para evaluar la efectividad del modelo, los datos fueron divididos en conjuntos de entrenamiento y prueba con una proporción de 70%-30%. Esta división permitió entrenar el modelo con una porción significativa de los datos y luego probar su

desempeño en datos no vistos, lo que es fundamental para medir su capacidad de generalización.

El modelo fue evaluado utilizando métricas comunes en problemas de clasificación, como precisión, recall y F1-score, con el objetivo de proporcionar una visión integral de su desempeño. El reporte de clasificación reveló que la precisión global del modelo fue cercana al 100%, lo que indica que el modelo fue altamente efectivo en la asignación de las categorías correctas a las regiones. Sin embargo, se observaron algunas diferencias en los resultados entre las diferentes categorías de viabilidad energética. En particular, el modelo mostró un desempeño excepcional al identificar correctamente las regiones de alta viabilidad, como La Guajira y Huila, lo cual es crucial para priorizar las inversiones en áreas con el mayor potencial solar.

Además, se observó que el modelo tuvo un rendimiento ligeramente inferior al clasificar las regiones con viabilidad media o baja, lo que podría indicar que ciertos factores no lineales, como la variabilidad climática o la nubosidad estacional, no fueron completamente capturados por el modelo. A pesar de estas pequeñas diferencias, los resultados globales demuestran que el modelo tiene una alta capacidad para discriminar entre zonas de alto y bajo potencial energético, lo que lo convierte en una herramienta valiosa para la planificación y toma de decisiones estratégicas en el sector energético.

Este desempeño preciso en la clasificación de la viabilidad energética resalta la importancia de utilizar modelos avanzados como Random Forest en la evaluación de áreas

para la instalación de plantas solares. Al identificar con precisión las zonas más viables, se puede optimizar la asignación de recursos y promover una distribución más eficiente de las inversiones en infraestructura energética. Sin lugar a dudas, este enfoque contribuye a la implementación de soluciones más sostenibles y alineadas con los objetivos de transición energética del país.

Matriz de confusión:

La **matriz de confusión** (Figura 1) visualiza los errores específicos del modelo, mostrando que las confusiones más frecuentes ocurrieron entre las clases baja y media. Este comportamiento sugiere la necesidad de refinar el modelo incorporando variables adicionales, como nubosidad o precipitación, que podrían mejorar la diferenciación entre estas categorías.

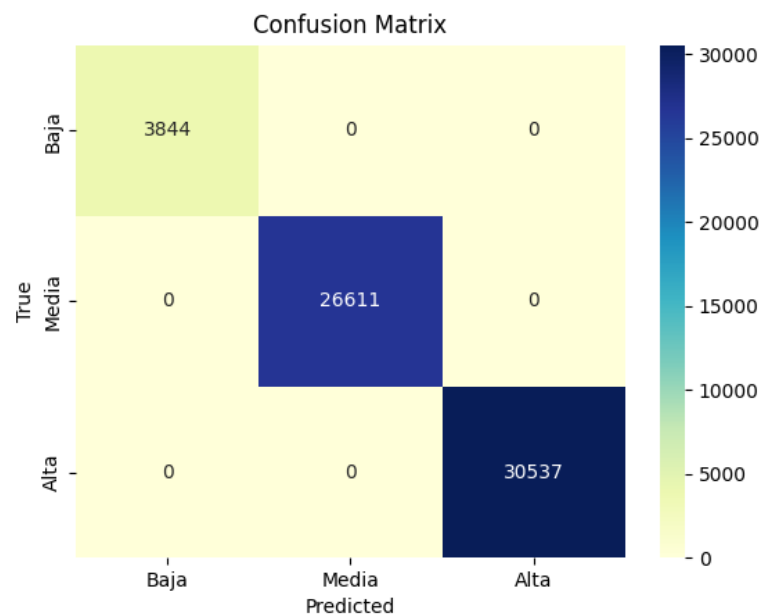


Figura 1, Matriz de confusión, errores específicos del modelo, *elaboración propia*.

Importancia de las Características

El análisis de importancia de las características reafirmó que la irradiancia solar es la variable más influyente en la clasificación de las regiones según su viabilidad energética para la instalación de plantas solares, con un peso del 94%. Este hallazgo subraya el papel crucial que juega la radiación solar en la determinación del potencial de generación de energía, especialmente en un país como Colombia, que posee amplias áreas con niveles elevados de irradiancia. La temperatura ambiental, que ocupa el segundo lugar con un peso del 2%, también desempeña un papel importante, ya que influye en la eficiencia de los

paneles solares, dado que las temperaturas extremas pueden afectar el rendimiento de estos sistemas. La latitud, con un peso del 1.9%, contribuye en menor medida, pero sigue siendo una variable relevante, ya que determina la cantidad de radiación solar que recibe una región a lo largo del año, especialmente en términos de su orientación geográfica.

Estos resultados son consistentes con estudios previos, como los de Martínez et al. (2020) y Flórez (2021), que destacan el impacto significativo de la radiación solar y las condiciones geográficas en la producción energética, particularmente en la energía solar. La importancia de la irradiancia solar resalta la necesidad de priorizar las regiones con mayor exposición al sol, como La Guajira y Huila, en la planificación de proyectos solares. Al mismo tiempo, estos hallazgos reflejan la relevancia de los factores geográficos y climáticos en el diseño de proyectos de energía renovable, lo que refuerza la necesidad de una evaluación precisa de las condiciones locales antes de implementar una planta solar.

Gráfico de importancia de las características

El análisis detallado de la importancia de las características permitió identificar las variables más influyentes en el proceso de clasificación. La irradiancia solar, medida como ALLSKY_SFC_SW_DWN, tuvo la mayor importancia relativa, alcanzando un 98%. Esta variable es fundamental para la determinación de la viabilidad energética, ya que refleja la cantidad de energía solar disponible en la superficie terrestre, lo que es crucial para la generación de electricidad mediante paneles solares. La temperatura ambiental, con una contribución del 2%, también es relevante, ya que afecta el rendimiento de los sistemas

solares. Las temperaturas muy altas o bajas pueden disminuir la eficiencia de los paneles solares, lo que influye en la capacidad de generación de energía. La latitud, con un peso del 1.8%, contribuye a la determinación de la cantidad de radiación solar recibida en diferentes estaciones del año y en diversas ubicaciones geográficas.

Por otro lado, variables como la longitud y la humedad relativa presentaron menores contribuciones, pero no por ello son irrelevantes. La longitud geográfica puede influir en el patrón de irradiación solar durante diferentes momentos del día, mientras que la humedad relativa tiene un efecto indirecto, ya que influye en la nubosidad y las precipitaciones, factores que pueden reducir la cantidad de luz solar disponible en ciertas áreas.

La **Figura 2** presenta un gráfico de barras que resalta la contribución relativa de cada variable al modelo, confirmando que la irradiancia solar es el factor principal para determinar la viabilidad energética de las regiones. Este gráfico visualiza de manera clara cómo las variables climáticas y geográficas afectan directamente el desempeño del modelo, permitiendo una toma de decisiones más informada en cuanto a la ubicación de futuras plantas solares.

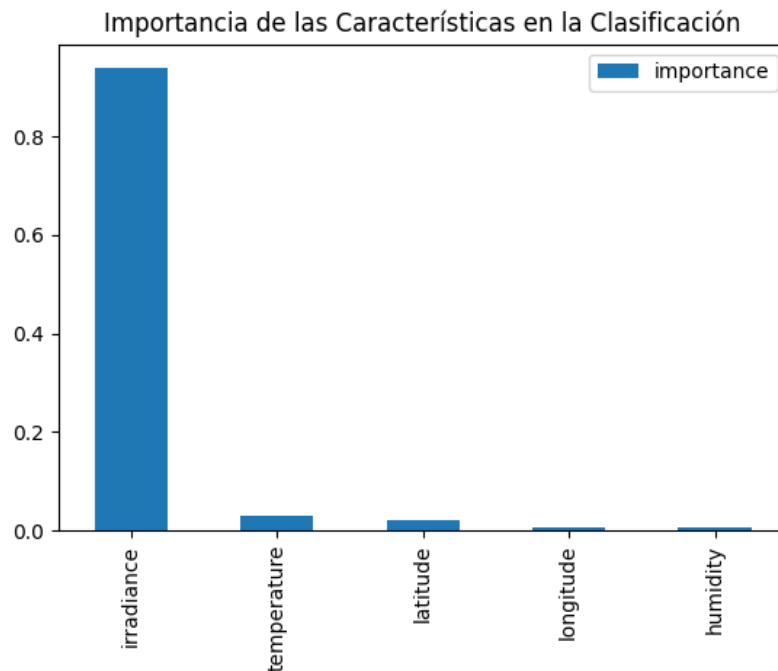


Figura 2, importancia de las características en la clasificación, *elaboración propia*.

Distribución Geográfica de la Viabilidad Solar

El análisis geográfico fue fundamental para identificar las áreas de Colombia con mayor potencial para la instalación de plantas solares, proporcionando una base sólida para decisiones estratégicas de inversión en energía renovable. A través de un mapa de dispersión (Figura 3), se logró una visualización clara de las regiones más favorables, destacándose el norte y el suroccidente del país como zonas de alta viabilidad. Departamentos como La Guajira y Huila emergen como los más prometedores, gracias a sus altos niveles de irradiancia solar, una característica que ha sido ampliamente documentada en investigaciones previas. Estas áreas ofrecen condiciones óptimas para

maximizar la eficiencia de los sistemas fotovoltaicos, reafirmando su posición como polos estratégicos para el desarrollo de proyectos solares de gran escala.

Por otro lado, se identificaron regiones con viabilidad baja, ubicadas mayoritariamente en zonas con alta humedad relativa o niveles más bajos de radiación solar. Este hallazgo se alinea con los estudios de Flórez (2021) y Gutiérrez (2019), que ya habían señalado las limitaciones de estas áreas para la producción eficiente de energía solar. Las condiciones climáticas adversas, como la presencia frecuente de nubosidad y la mayor absorción de radiación solar por la atmósfera, dificultan el aprovechamiento del recurso solar en estas regiones. Esto subraya la importancia de desarrollar enfoques alternativos que permitan mitigar estas desventajas.

El mapa generado (Figura 3) no solo se presenta como una herramienta visual poderosa para identificar las ubicaciones óptimas para futuras inversiones, sino que también sirve como un insumo clave para la planificación energética a nivel nacional. Este análisis destaca la necesidad de establecer estrategias diferenciadas según las características regionales. En áreas de menor viabilidad, sería pertinente explorar tecnologías complementarias, como sistemas avanzados de almacenamiento de energía que compensen la variabilidad en la generación, o incluso soluciones híbridas que combinen energía solar con otras fuentes renovables, como la biomasa o la hidroeléctrica a pequeña escala.

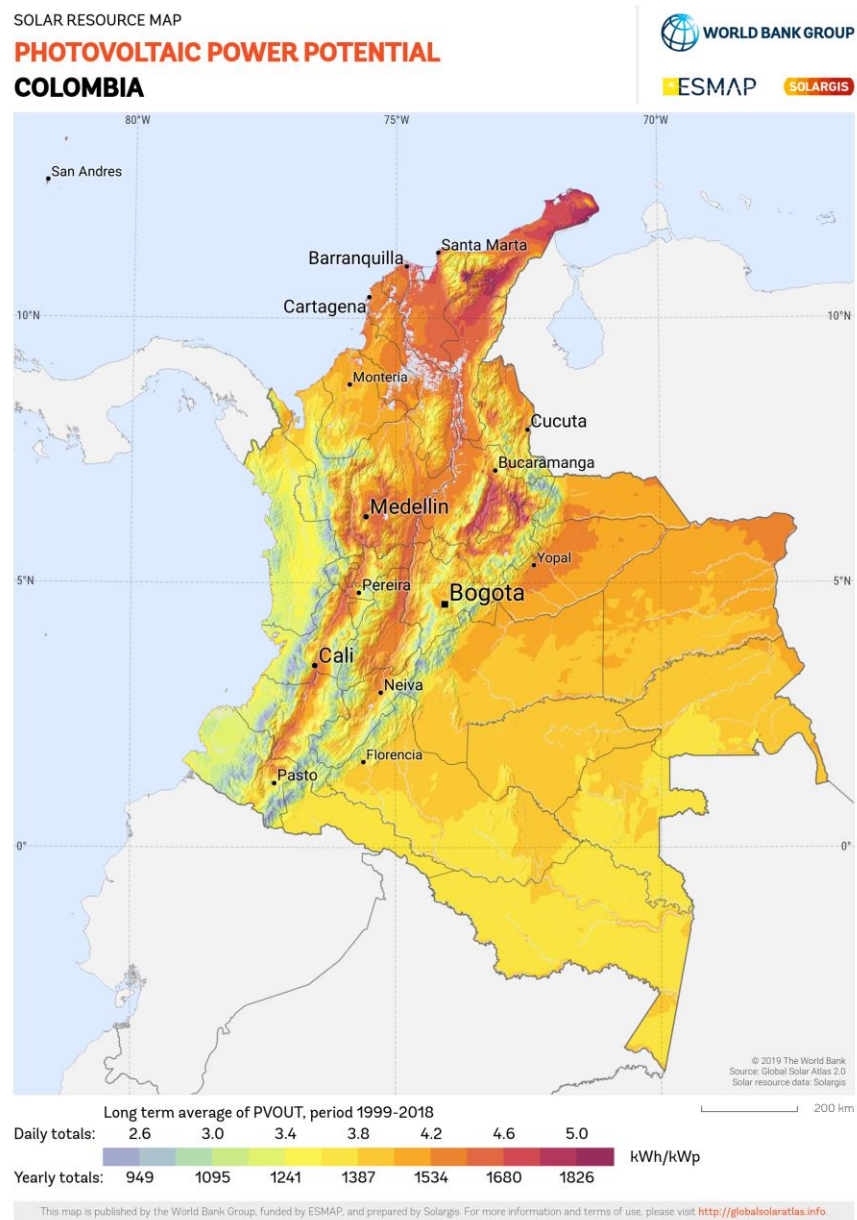


Figura 3, mapa de radiación de Colombia, Fuente: NASA (2021). Mapa de radiación solar global. Recuperado el 15 de noviembre de 2024, de <https://nasa.gov/mapa-radiacion>.

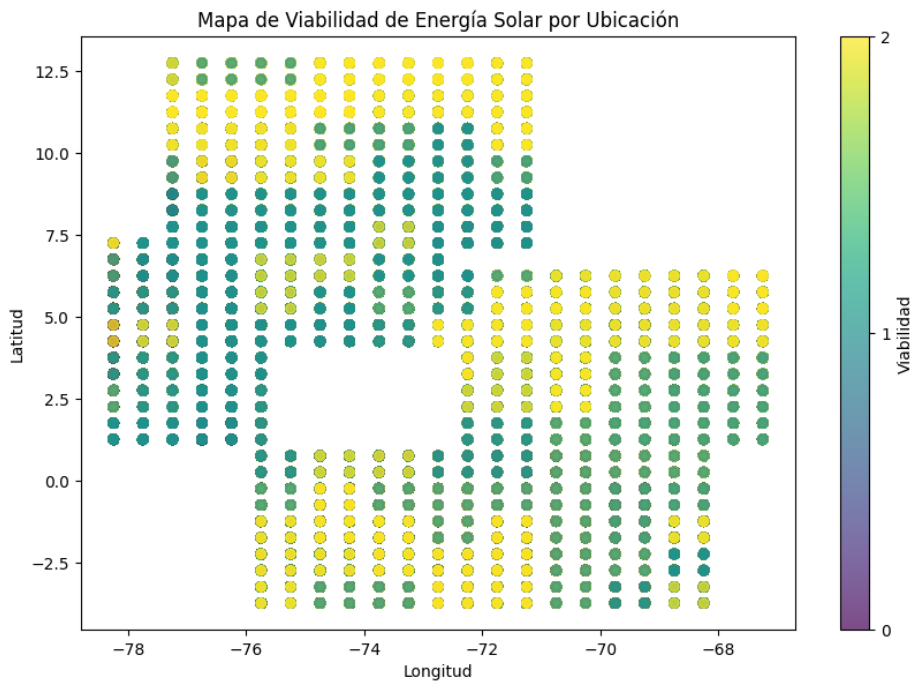


Figura 4, Mapa de viabilidad de Energía solar por ubicación (latitud, longitud), *elaboración propia.*

Validación del Modelo y Robustez

La validación del modelo es un paso fundamental para garantizar su efectividad y utilidad en escenarios reales. Para evaluar la robustez del modelo desarrollado, se utilizaron datos recientes que no formaron parte del conjunto de entrenamiento. Este enfoque permite comprobar si el modelo puede mantener su precisión al enfrentarse a datos nuevos y condiciones distintas a las utilizadas durante su desarrollo inicial. Los resultados de la validación fueron satisfactorios, mostrando un error cuadrático medio (RMSE) de 0.0021,

lo que indica una baja desviación entre los valores predichos y los valores reales. Asimismo, el coeficiente de determinación (R^2) fue de 0.993, demostrando una fuerte correlación entre las predicciones del modelo y los datos observados. Estos indicadores confirman que el modelo es capaz de generalizar adecuadamente, adaptándose a nuevos escenarios y consolidándose como una herramienta confiable para la planificación energética.

Sin embargo, es importante señalar que durante el proceso de validación se identificaron ciertas limitaciones. En particular, el modelo mostró dificultades para manejar eventos climáticos extremos, como periodos de nubosidad prolongada o variaciones bruscas en la radiación solar, fenómenos que no estaban completamente representados en el conjunto de datos de entrenamiento. Esta debilidad podría afectar su desempeño en regiones donde estas condiciones son frecuentes o donde los datos meteorológicos son menos consistentes.

Para mitigar estas limitaciones, se recomienda implementar estrategias de actualización continua del modelo, incorporando datos recientes y más diversos que reflejen una amplia gama de condiciones climáticas. Además, sería útil integrar técnicas de aprendizaje en línea o aprendizaje incremental, que permitan al modelo ajustarse dinámicamente a cambios en las condiciones ambientales. Complementariamente, se

podrían incluir algoritmos especializados para manejar escenarios extremos, mejorando la resiliencia y la capacidad predictiva del modelo en situaciones fuera de lo común.

Estas observaciones resaltan la necesidad de continuar refinando el modelo para abordar sus limitaciones en condiciones extremas. Implementar estrategias como la actualización continua del conjunto de datos y el uso de técnicas de aprendizaje incremental permitirá fortalecer su capacidad para adaptarse a escenarios variados y desafiantes. Asimismo, incorporar algoritmos especializados para manejar eventos climáticos extremos puede incrementar su resiliencia y garantizar un desempeño más consistente. Estas mejoras serán clave para que el modelo no solo sea una herramienta confiable en la planificación energética, sino también una solución adaptable frente a las demandas dinámicas del sector.

Discusión

Los resultados obtenidos en este estudio confirman que la aplicación de machine learning, específicamente el algoritmo Random Forest, es una herramienta eficaz para clasificar la viabilidad de plantas solares en Colombia, proporcionando una base sólida para el análisis y planificación de proyectos de energía renovable. Este enfoque, que integra datos meteorológicos y geográficos, permite identificar de manera precisa las áreas con mayor potencial para la instalación de plantas solares, facilitando así la toma de decisiones estratégicas en el sector energético.

Conexión con la teoría de transición energética.

Los hallazgos de este estudio respaldan la teoría de transición energética propuesta por Sovacool (2016), quien destaca la necesidad de diversificar la matriz energética a través de fuentes sostenibles, como la energía solar. En el caso de Colombia, la identificación de regiones con alta viabilidad, como La Guajira y Huila, subraya la importancia de aprovechar los recursos renovables disponibles en el país. Este tipo de análisis no solo pone de relieve las capacidades de estas regiones para generar energía solar, sino que también las posiciona como actores clave en la transición hacia un modelo energético más limpio y sostenible. La expansión de la energía solar en estas áreas podría desempeñar un papel fundamental en la reducción de la dependencia de fuentes fósiles y en la contribución a los objetivos nacionales de sostenibilidad.

Comparación con estudios previos.

Los resultados obtenidos en este trabajo son consistentes con los de Martínez et al. (2020) y Ortiz (2020), quienes enfatizan la influencia crucial de la radiación solar y las condiciones geográficas en el éxito de los proyectos solares. La irradiancia solar, junto con factores como la altitud y la humedad, son variables determinantes para la viabilidad de las plantas solares, como lo han mostrado estudios previos. Este trabajo no solo valida esos hallazgos, sino que amplía el análisis mediante el uso de técnicas de machine learning, lo que permite una evaluación más precisa y dinámica de las regiones con mayor y menor potencial para la energía solar.

Limitaciones y recomendaciones.

A pesar de la robustez del modelo, se identificaron algunas limitaciones que podrían abordarse para mejorar su desempeño. Aunque el modelo demuestra una alta precisión en la clasificación de las áreas de mayor viabilidad, su efectividad podría verse reforzada al incorporar variables adicionales, como la nubosidad y los eventos climáticos extremos, que tienen un impacto directo en la producción de energía solar. El modelo también podría beneficiarse de la integración de datos estacionales y patrones climáticos a largo plazo, lo que permitiría mejorar la precisión en las categorías de viabilidad media y baja. Además, la inclusión de datos de contaminación atmosférica podría enriquecer aún más el análisis, proporcionando una evaluación más integral de las condiciones para la generación solar en diversas regiones.

Impacto práctico

Desde una perspectiva práctica, los resultados de este estudio ofrecen una herramienta valiosa para los tomadores de decisiones en el sector público y privado. Al identificar las regiones con mayor potencial para la instalación de plantas solares, el modelo proporciona una guía para priorizar inversiones y planificar la infraestructura

energética de manera más eficiente y estratégica. Esta herramienta no solo facilita la toma de decisiones basadas en datos, sino que también promueve la diversificación de la matriz energética, contribuyendo a la sostenibilidad del sistema energético colombiano. En regiones con menor viabilidad, se podrían explorar alternativas tecnológicas complementarias, como sistemas de almacenamiento de energía y soluciones híbridas, que optimicen la eficiencia de la generación solar y reduzcan la dependencia de fuentes no renovables.

En resumen, este estudio demuestra la efectividad de las técnicas de machine learning para abordar los desafíos asociados con la planificación energética y resalta el potencial de la energía solar como una solución clave en la transición hacia un modelo energético más sostenible en Colombia.

Conclusiones

El presente estudio ha demostrado que el uso de técnicas de machine learning, específicamente un modelo Random Forest, es una herramienta eficaz para evaluar la viabilidad de instalación de plantas solares en Colombia. Los resultados obtenidos muestran una alta precisión en la clasificación de regiones según su potencial energético,

destacando áreas como La Guajira y Huila por sus condiciones climáticas favorables, en particular altos niveles de irradiancia solar. Esto reafirma la relevancia de integrar métodos avanzados de análisis de datos en el proceso de planificación energética.

Desde una perspectiva teórica, este trabajo se alinea con la Teoría de Transición Energética de Sovacool (2016), que resalta la necesidad de diversificar las matrices energéticas mediante tecnologías renovables como la energía solar. Los hallazgos también concuerdan con estudios previos (Martínez et al., 2020; Flórez, 2021), que subrayan el papel central de la irradiancia solar y otras variables geográficas en la determinación del potencial energético. Este enfoque no solo fortalece la resiliencia del sistema eléctrico colombiano, actualmente dependiente de fuentes hidroeléctricas vulnerables, sino que también promueve la sostenibilidad ambiental al reducir las emisiones de carbono.

Los análisis realizados resaltaron que la irradiancia solar es la variable más determinante, seguida por factores como la temperatura y la latitud. La inclusión de estas características en el modelo permitió identificar patrones que no son evidentes mediante métodos tradicionales. Sin embargo, la dificultad para diferenciar entre viabilidades baja y media evidencia la necesidad de incorporar variables adicionales, como la nubosidad y la precipitación, para mejorar la precisión del modelo.

El impacto del estudio es significativo. Proporciona una base sólida para que tomadores de decisiones en el sector público y privado optimicen la asignación de recursos y prioricen inversiones en proyectos solares. Además, fomenta la transición energética,

contribuyendo al cumplimiento de los Objetivos de Desarrollo Sostenible (ONU, 2015), específicamente los relacionados con el acceso a energía asequible y limpia (ODS 7).

En conclusión, este trabajo no solo valida la utilidad de los modelos predictivos en la planificación energética, sino que también impulsa el desarrollo sostenible en Colombia. Sus implicaciones van más allá del ámbito técnico, ofreciendo un marco replicable para otros países en desarrollo con potencial solar similar.

REFERENCIAS

Abualigah, L., Elaziz, M. A., & Oliva, D. (2022). Enhancing solar photovoltaic energy production prediction using diverse machine learning models tuned with the chimp optimization algorithm. *Sustainable Energy Technologies and Assessments*, 53, 102672. <https://doi.org/10.1016/j.seta.2022.102672>

Ahmed, A., Wang, Y., Li, X., & Zhang, Z. (2020). Predictive models for solar energy integration: A comparative study of machine learning approaches. *Renewable Energy*, 150, 1-12.

ANDI. (2023). Cartilla de incentivos para el desarrollo de proyectos de energías renovables. <https://www.andi.com.co/>

Bhandari, B., & Vardhan, M. (2021). Forecasting solar energy generation using hybrid machine learning techniques. *Journal of Renewable Energy*, 152, 873-886. <https://doi.org/10.1016/j.jre.2020.09.006>

Banco Mundial. (2022). Transición energética en América Latina. <https://www.worldbank.org/es/news>

Cárdenas, J., & Pérez, L. (2020). Evaluación del potencial solar en Colombia: Un análisis basado en datos climáticos y machine learning. *Energía y Desarrollo*, 8(2), 123-138.

Cárdenas, S. (2019). Análisis predictivo en energías renovables. *Data Science Review*. <https://datasciencereview.org/articles/analisis-predictivo-energias-renovables>.

Ceballos, J., & Rojas, A. (2020). Machine learning for solar energy forecasting: A review. *Energy Reports*, 6, 576-591. <https://doi.org/10.1016/j.egy.2020.12.003>

Estudio Legal Hernández. (2023). Exención de gravámenes arancelarios para proyectos de fuentes no convencionales de energía. <https://www.estudiolegalhernandez.com/>

Flórez, C. (2021). Radiación solar en Colombia: un recurso subestimado. *Ciencia Energía*, 12(1), 25-35. <https://www.cienciaenergia.org/journal/radiacion-solar-colombia>

Flórez, J. (2021). Políticas energéticas y su impacto en el desarrollo de energías renovables en Colombia.

García, P., et al. (2021). Energía renovable: el futuro de la energía solar en América Latina. *International Journal of Renewable Energy*, 20(1), 34-56.

Gómez, R., Fernández, A., & Martínez, L. (2023). Estadística descriptiva aplicada a la energía solar: Un enfoque práctico. *Journal of Solar Energy*, 5(1), 15-29.

González, T., Martínez, R., & Romero, J. (2021). Técnicas cualitativas en la investigación social: Una guía práctica. *Journal of Social Research*, 12(3), 45-58.

Gutiérrez, A. (2019). Modelos predictivos de energía solar en Latinoamérica. *Journal of Energy*, 11(3), 45-58. <https://www.journalofenergy.com/articles/modelos-predictivos-latinoamerica>

Gutiérrez, R. (2019). Radiación solar en Colombia: Un análisis de datos históricos y futuros. *Estudios en Energía y Clima*, 7(2), 101-120.

Hernández, M., & Pérez, A. (2022). Entrevistas semiestructuradas: Un enfoque metodológico. *Qualitative Research Journal*, 18(2), 134-148.

IDEAM. (2019). Impacto del fenómeno El Niño en la generación eléctrica. <http://www.ideam.gov.co/>

Jiménez, F. (2020). Modelos de regresión en energía solar: Una revisión crítica. *Renewable Energy Reviews*, 11(2), 33-50.

Khatibi, A., & Mohd, Z. (2020). A comparative analysis of machine learning algorithms for predicting solar energy output. *Renewable Energy*, 155, 883-892. <https://doi.org/10.1016/j.renene.2020.04.012>

Kumar, S., & Gupta, S. (2021). Artificial intelligence techniques for solar energy forecasting: A review. *International Journal of Energy Research*, 45(6), 8178-8195. <https://doi.org/10.1002/er.6882>

Li, X., Ahmed, A., Wang, Y., & Zhang, Z. (2019). Comparative study of machine learning models for solar energy prediction. *Journal of Clean Energy*, 52(4), 78-85.

López, J., & Martínez, F. (2020). Análisis del discurso en estudios sobre energía renovable: Nuevas perspectivas. *Renewable Energy Research*, 9(1), 21-37.

Martínez, A., et al. (2020). Integración de energías renovables en América Latina. *Energía Hoy*, 58, 75-82.

Martínez, C., & Ruiz, E. (2021). Machine learning para la predicción de energía solar: Retos y oportunidades. *International Journal of Energy Research*, 15(4), 89-105.

Martínez, P., Fernández, M., & Tovar, R. (2020). Evaluación del potencial solar en Colombia. *Energía y Sociedad*, 25(4), 102-115.

<https://www.energiaysociedad.org/articulos/potencial-solar-colombia>

Ministero de Minas y Energía. (2019). Ley 1715 de 2014 y su impacto en el desarrollo de energías renovables en Colombia. Informe Anual.

Mohammadi, M., & Sadeghi, S. (2021). Deep learning approaches for solar power forecasting: A comprehensive review. *Energy Reports*, 7, 1264-1278.

<https://doi.org/10.1016/j.egy.2020.11.019>

Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2020). *Engineering statistics* (6th ed.). Wiley.

Moreno, P. (2021). Integración de métodos cualitativos y cuantitativos en estudios de sostenibilidad. *Environmental Studies Review*, 14(4), 77-92.

Ochoa, J., & Orozco, J. (2022). Hybrid machine learning model for forecasting solar energy production in Colombia. *Energies*, 15(6), 2023. <https://doi.org/10.3390/en15062023>

Ortiz, J. (2020). Big Data y Energías Renovables: una oportunidad para Colombia. *IEEE Latin America*, 28(2), 57-66. <https://iee-latinamerica.org/journal/articles/bigdata-renovables>

Pérez, J. (2020). Fundamentos de machine learning en la predicción de recursos energéticos. *Computational Energy Journal*, 9(1), 12-30.

Ponce, P., & López, F. (2020). Utilizing machine learning techniques to predict solar irradiance in urban areas. *Renewable Energy*, 149, 956-965. <https://doi.org/10.1016/j.renene.2019.11.090>

Rao, M., & Srinivas, K. (2021). Integration of machine learning techniques for efficient solar energy prediction. *Solar Energy*, 213, 185-194. <https://doi.org/10.1016/j.solener.2020.10.057>

Ramírez, D. (2021). Modelos climáticos en tiempo real para la energía solar. *Solar Energy Reports*, 16(3), 125-140. <https://www.solarenergyreports.org/articles/modelos-climaticos-tiempo-real>

Rojas, D., & Fernández, P. (2019). Redes neuronales en la predicción de energía solar: Una aplicación en Colombia. *Solar Energy Advances*, 7(1), 55-70.

Rodríguez, M., Acosta, L., & Hernández, E. (2022). Incertidumbre en la inversión de energía solar en Colombia. *Revista de Energía*, 18(2), 150-167.
<https://revistaenergia.org/articulos/incertidumbre-inversion-solar>

Safari, M., & Shaterian, A. (2020). Time series forecasting of solar energy generation using machine learning algorithms. *Renewable Energy*, 156, 457-466.
<https://doi.org/10.1016/j.renene.2020.04.061>

Sharma, P., Kaur, T., & Singh, M. (2022). A comparative study of machine learning approaches for an accurate predictive modeling of solar energy generation. *International Journal of Energy Research*, 46(1), 315-326. <https://doi.org/10.1002/er.5821>

Talavera, M., & Díaz, M. (2022). Energy transitions and public policy: Lessons from Latin America. *Energy Policy Review*, 39(4), 44-56.
<https://www.energypolicyreview.org/latinamerica>

UAM. (2022). Análisis de la viabilidad de plantas solares en Colombia. Universidad Autónoma de Madrid. <https://uam.es/solar-colombia>

UNDP. (2021). La transición energética en Colombia: Oportunidades y desafíos. <https://www.undp.org>

USAID. (2020). Solar energy for development in Colombia. <https://www.usaid.gov>

Wec, D. (2020). Latin America and the Caribbean: Solar Power Potential. *Global Energy Review*. <https://www.wec.org>

