

# Uso de inteligencia artificial para mejorar el desempeño de un proyecto de generación fotovoltaica

Seminario de Analítica y Ciencia de Datos

MOISÉS GUERRERO JIMÉNEZ, ANDRÉS CASTAÑO LICONA

November 4, 2023

## 1 Resumen Descriptivo del Proyecto

Este proyecto tiene como objetivo abordar el desafío de aportar en la optimización de la gestión de energía solar fotovoltaica a partir del desarrollo de modelos que permitan predecir la disponibilidad de radiación solar en la ubicación del futuro Parque Solar Fotovoltaico Tepuy en La Dorada, departamento de Caldas. Haciendo uso de técnicas de aprendizaje automático sobre datos de irradiancia solar e información meteorológica medidos a lo largo del tiempo en la ubicación del proyecto.

### 1.1 Problema

La radiación solar es un término que hace referencia a la radiación electromagnética emitida por el sol. Esta puede ser capturada y convertida en formas útiles de energía tales como calor y electricidad por medio de diferentes tecnologías. Sin embargo, la viabilidad del uso de estas tanto a nivel técnico como a nivel económico dependen en gran medida de la disponibilidad de radiación solar [1]. En el caso de las celdas solares como las usadas en el proyecto Tepuy, se ha encontrado que una baja radiación solar tiene un impacto significativo en la calidad de la potencia eléctrica generada por el sistema fotovoltaico [2].

Con el fin de optimizar la calidad de energía recibida, es posible realizar ajustes de posicionamiento de las celdas solares de acuerdo a la disponibilidad de radiación solar [3], por lo que realizar predicciones de la radiación solar permitiría ayudar a optimizar la calidad de la potencia eléctrica generada a partir de un posicionamiento oportuno de las celdas solares.

De acuerdo a [1], todos los lugares en la tierra reciben luz solar al menos una parte del año, puesto que la cantidad de radiación solar que llega a cualquier punto en la superficie terrestre, dependen de la ubicación geográfica, la hora del día, la estación del año, el paisaje y la climatología local.

A medida que la luz solar atraviesa la atmósfera, parte de ella es absorbida, dispersada y reflejada por las moléculas de aire, el vapor de agua, las nubes o el polvo, entre otros.

Según lo indicado en [3] la radiación total en la superficie terrestre procedente del sol se puede dividir en tres partes principales: radiación directa, difusa y reflejada. En primer lugar, la radiación directa se ve afectada por la cantidad de vapor de agua, la densidad atmosférica y la concentración de emisiones en el aire. En segundo lugar, la radiación difusa es la que llega a la superficie terrestre desde el cielo, no directamente desde el sol sino como luz dispersada y reflejada por la composición de la atmósfera. En consecuencia, la radiación difusa es difícil de estimar ya que se ve muy afectada por la nubosidad que es muy variable.

De acuerdo a [4], el pronóstico a corto plazo sobre la producción de energía esperada puede permitir una integración eficiente de las fuentes de energía renovables a través del comercio eficiente de energía, el control del sistema eléctrico y la gestión de unidades de almacenamiento de energía eléctrica. Esto es clave para los equipos encargados del diseño y optimización de los sistemas fotovoltaicos que se encuentran en instalación por parte de EPM en la ubicación del Parque Solar Fotovoltaico Tepuy, donde se contará con un total aproximado de 200.000 celdas solares. Ya que, al poder realizar predicciones de la radiación solar disponible de acuerdo a ciertos factores ambientales, se puede apoyar en una mejor gestión de la energía eléctrica generada para su conexión con la red eléctrica nacional.

## 1.2 Datos

Se realizó un acercamiento con un funcionario de EPM perteneciente a la Unidad de Hidrometría y Calidad Generación, que es el área responsable del gobierno de los datos del proyecto Tepuy y otros relativos a la generación de energía en EPM.

Según dicho funcionario, la fuente de información corresponde a datos medidos en una estación meteorológica ubicada en la zona de actual instalación del proyecto Tepuy en el municipio de La Dorada, Caldas. Las variables que nos fueron informadas, pero de las cuales no se ha hecho validación, son irradiación solar, temperatura, humedad relativa, precipitación, velocidad y dirección del viento, con frecuencia diezminutal. La serie de registros consta de aproximadamente 190.000 eventos medidos, desde el año 2020 a la fecha. Los sistemas de asimilación de datos son redundantes y hay validación de calidad de los mismos.

Elevamos una solicitud al jefe del área correspondiente de EPM para acceder a los datos para el proyecto de la materia de Seminario. La respuesta nos indica que es viable, si se atienden los siguientes requerimientos por parte de EPM:

- Para dejar evidencia de que es un trabajo con la Universidad de Antioquia, se solicita que una vez aprobado el anteproyecto, se comparta un comunicado desde la Universidad ratificando que será nuestro proyecto asignado.
- Antes de la entrega de la información, será necesario firmar un documento de manejo de información confidencial que será compartido una vez se formalice el proyecto con la Universidad.
- Dado que son datos de EPM, no se pueden incluir como anexos en el trabajo final ni publicar los mismos.
- El resultado del trabajo puede ser del interés de EPM, por lo cual se hace necesario que estemos en la disposición de darles a conocer los resultados y eventualmente permitirles utilizarlos para la operación, si aplica.

En la actualidad adelantamos el cumplimiento de estos requisitos para acceder a los datos a través del programa de la Especialización.

## 1.3 Hipótesis

La crisis del cambio climático ha generado que la comunidad científica destine grandes esfuerzos a encontrar alternativas a los combustibles fósiles, que es una de las principales fuentes del calentamiento global [5] y una de las principales fuentes de generación de energía en las centrales termoelectricas. La energía solar es entonces considerada como una fuente de energía alternativa muy prometedora para la transición energética. No obstante, también tiene importantes retos para integrar de manera confiable y estable al sistema de distribución [6]. Para cualquier sistema de generación de energía debe existir un correcto balance entre la demanda y la oferta. Las predicciones de estos dos componentes permiten a los integrantes del sistema (generadores, distribuidores, operadores, reguladores, entre otros) establecer las

dinámicas del mercado energético, como la compra y venta de energía. En ese sentido, para EPM, que actúa tanto como generador y distribuidor de energía, es completamente relevante establecer en el corto, mediano y largo plazo su capacidad de generación para ofrecer energía en firme y determinar si debe comprar energía para cumplir con sus obligaciones. Por tanto, grandes desviaciones en las predicciones tienen un alto costo en una empresa como EPM [7]. De tal manera, es supremamente deseable poder predecir la radiación solar, en diferentes escalas temporales, para un mercado como el colombiano que está realizando inversiones importantes en los sistemas de generación fotovoltaica a pesar de su complejidad, asociada a la variabilidad climática [8]. Estas predicciones tienen amplias aplicaciones: determinar los mejores lugares de instalación de parques solares, determinar el balance entre oferta y demanda, establecer la franja intra horaria para el despacho de energía al Sistema Interconectado Nacional (SIN), administrar el almacenamiento de energía, planificación de los ciclos de mantenimiento y planificar las transacciones en el mercado energético [9][10].

La radiación teórica depende esencialmente de aspectos espacio temporales y astronómicos como la latitud, longitud, día del año, hora, declinación solar, la distancia media de la tierra al sol y distancia de la tierra al sol para una fecha concreta [11]. No obstante, la radiación solar real, medida en tierra, depende de otros aspectos meteorológicos difíciles de predecir como la temperatura, la velocidad del viento, la humedad relativa y, claro está, la radiación teórica.

Nuestra hipótesis es que a través de modelos de inteligencia artificial se puede pronosticar la radiación solar del proyecto Tepuy en diferentes escalas temporales, basados en los datos de los que dispone EPM, capturando la variabilidad climática de la zona.

Las escalas temporales que pretendemos modelar y predecir son:

- **Escala de corto plazo:** en la franja diezminutal hasta 1-5 días. Ideal para aplicaciones relativas al mercado energético de compra y venta de energía y el balance entre oferta y demanda.
- **Escala de mediano:** en la franja de 5 días hasta 1-2 meses. Con este tipo de predicción se pueden programar mantenimientos en el sistema para mejorar la eficiencia y extender la vida útil de los equipos.

En la evaluación del modelo se puede determinar si las predicciones son cercanas a los valores reales, en las diferentes escalas. Para lo cual contemplamos tres posibles escenarios que nos llevarían a tomar acciones frente a los resultados obtenidos:

1. Los resultados del modelo son óptimos en alguna de las escalas temporales y se pueden implementar como una solución para el cliente y los stakeholders.
2. Los resultados del modelo no son suficientemente óptimos, para alguna de las escalas, y se requiere una mejor ajuste del modelo o la implementación de un modelo diferente, con su respectiva validación.
3. Los resultados obtenidos no capturan la variabilidad de la radiación en alguna de las escalas, aún después de explorar diferentes técnicas de ajuste o diferentes modelos. Por lo tanto, no existe forma de implementar la solución con el cliente.

## 1.4 Solución

La energía eléctrica generada a partir de fuentes de energía renovables presenta cierta inestabilidad en el tiempo, debido a los factores ambientales previamente revisados. Esta inestabilidad impacta en la capacidad de predicción, algo que se ve reforzado al entrenar con datos en intervalos de minutos como lo es nuestro caso, haciendo necesario adoptar metodologías sofisticadas para la previsión de la producción de los sistemas energéticos [12].

Tal como se menciona en [12], los modelos que presentan un mejor rendimiento predictivo en comparación, para la radiación solar son los modelos de aprendizaje profundo (ANN, CNN y LSTM), los modelos híbridos de aprendizaje profundo (CNN-ANN y CNN-LSTM-ANN) y puntualmente el modelo XGB (Extreme Gradient Boosting). Sin embargo, de acuerdo a los métodos observados en [13] y los resultados obtenidos por métodos como GBRT, SVM, Decision Tree Regressor, Random Forest o RLM, se realizarán implementaciones con estos modelos, siguiendo un enfoque de exploración y comparación, evaluando el rendimiento de algunos de estos y su capacidad para realizar predicciones con los datos en cuestión. Esta implementación de múltiples modelos, también permitirá seguir un enfoque de aprendizaje conjunto entre varios de estos modelos.

Para la implementación de la solución se debe tener en cuenta los siguientes aspectos:

- La implementación sobre el conjunto de datos se enfocará inicialmente en el análisis estadístico con el fin de identificar dependencias entre predictores usados.
- El tratamiento de los datos y el abordaje de los modelos debe ser orientado a series de tiempo, donde los datos son dependientes de los valores inmediatamente anteriores y los muestreos no son aleatorios.
- Posteriormente, se seguirá un enfoque de aprendizaje conjunto presentado en [13], integrando predicciones de varios modelos, permitiendo combinar las fortalezas de diferentes enfoques de modelado para mejorar la precisión de las predicciones.

Los resultados esperados de la solución es que se puedan implementar varios modelos que permitan evaluar su desempeño y escoger los mejores, atendiendo los siguientes aspectos:

- **Evaluación de modelos:** Los modelos desarrollados serán evaluados para determinar su precisión y capacidad de predicción. Se compararán las predicciones generadas por los modelos con los valores reales de radiación solar, utilizando métricas de evaluación adecuadas, descritas más adelante. Esta evaluación permitirá identificar qué modelos ofrecen un rendimiento óptimo en las diferentes escalas temporales y cuáles requieren ajustes adicionales.
- **Selección de modelos óptimos:** Basándonos en los resultados de la evaluación, se seleccionarán los modelos que demuestren un rendimiento sobresaliente en la predicción de la radiación solar. Estos modelos se considerarán soluciones potenciales para mejorar la gestión de energía solar en el Parque Solar Fotovoltaico Tepuy. La selección se basará en la capacidad de los modelos para proporcionar predicciones precisas y confiables, en las escalas temporales del mediano plazo y/o mediano plazo, que respalden la toma de decisiones efectivas en la operación del parque.

## 1.5 KPIs

La evaluación del desempeño del modelo se realizará según las escalas temporales definidas, a corto y mediano plazo. En ambos casos, una vez definida la frecuencia de la escala, se comparará el valor de la predicción con el dato real.

**Validación del modelo para el corto plazo:** en vista de que los datos de entrada tienen frecuencia diezminutal, se espera obtener datos de salida para el corto plazo con la misma frecuencia. Para ello se deberá escoger una ventana de tiempo que servirá como el conjunto de pruebas (entre uno y dos días), en los que haya condiciones climáticas distintas (temporada seca y temporada de lluvias, por ejemplo).

**Validación del modelo para el mediano plazo:** en esta escala de tiempo deberá seleccionarse la frecuencia óptima (entre 5 días a uno o dos meses). Los datos reales deberán escalarse a la frecuencia seleccionada para el mediano plazo (mediante alguna técnica de agregación) y compararse con los valores predichos por el modelo. Se espera que los resultados en el mediano plazo tengan una mejor evaluación que los datos predichos para el corto plazo, donde hay una mayor variabilidad intra horaria.

### 1.5.1 Métricas

Las métricas que se usarán son las estándar para los modelos de machine learning, según lo refiere la literatura [11], a saber:

**Error absoluto medio (MAE):** Se calcula como el promedio de la diferencia absoluta entre el valor real de la radiación solar y el valor predicho. Esta métrica es útil para penalizar el modelo tanto si realiza sobreestimaciones como subestimaciones.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (1)$$

Donde  $y_i$  es el vaor de radiación solar,  $\hat{y}_i$  es el valor de radiación predicho y  $N$  es el número total de muestras.

**Error cuadrático medio (MSE):** Se calcula como el promedio del cuadrado de las diferencias entre el valor de radiación predicho y el valor real. Esta métrica penaliza los casos en los que las diferencias entre ambos valores son mayores.

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (2)$$

**Raíz del error cuadrático medio (RMSE):** Se calcula como la raíz cuadrada del promedio del cuadrado de las diferencias entre el valor de radiación predicho y el valor real. Esta métrica suele ser una de las más utilizadas para evaluar el desempeño de un modelo. También ayuda a identificar y eliminar outliers en los datos.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

## 1.6 Actores

A continuación se listan los actores que identificamos hacen parte de la solución propuesta: cliente, interesados (stakeholders), usuarios e impactados.

### 1.6.1 Cliente

EPM es la entidad responsable del Parque Solar Fotovoltaico Tepuy y, por lo tanto, sería el principal receptor de los resultados y soluciones desarrolladas en este proyecto, actuando como cliente principal que utilizará la solución para tomar decisiones a partir de los resultados obtenidos y de esta manera mejorar la operación del parque solar.

### 1.6.2 Interesados (Stackeholders)

- **EPM (Empresas Públicas de Medellín):** Como la entidad responsable del proyecto y la generación de energía, EPM es el principal stakeholder. Están interesados en maximizar la eficiencia de la generación de energía solar en el Parque Solar Fotovoltaico Tepuy, reducir costos y optimizar la toma de decisiones relacionadas con la energía solar.

- **Inversionistas:** Cualquier entidad o individuo que haya invertido en el proyecto tiene un interés significativo en su éxito, ya que afectará el retorno de la inversión y la rentabilidad.
- **Autoridades reguladoras:** Las agencias gubernamentales encargadas de la regulación y supervisión de la generación de energía, así como de las prácticas medioambientales, pueden ser stakeholders importantes.
- **Medio ambiente:** Dado que el proyecto implica el uso de energía solar, es fundamental tener en cuenta las cuestiones ambientales. Los stakeholders ambientales se preocupan por la sostenibilidad y la reducción de las emisiones de carbono.
- **Clientes y consumidores de energía:** Los clientes que recibirán la energía generada por el parque solar también son stakeholders, ya que desean una fuente confiable y eficiente de energía.

### 1.6.3 Usuarios de la solución

- Los gestores y analistas de energía en EPM utilizarán la solución para optimizar la generación y distribución de energía solar en el contexto de la red eléctrica nacional. Esto incluye la planificación de la compra y venta de energía en el mercado energético.
- El personal encargado de la operación y mantenimiento del parque solar utilizará la solución para tomar decisiones sobre el posicionamiento de las celdas solares, la programación de mantenimiento y la gestión de la energía generada.

### 1.6.4 Impactados

- EPM se beneficiará directamente al lograr una gestión más eficiente de la energía generada por el Parque Solar Fotovoltaico Tepuy. Esto puede traducirse en una mayor rentabilidad, una generación de energía más confiable y una mejor planificación operativa.
- La implementación exitosa de esta solución podría tener un impacto positivo en el sector energético de Colombia al promover la generación de energía solar y contribuir a la transición hacia fuentes de energía más limpias y sostenibles.

## Referencias

1. U.S. Department of Energy. Solar Radiation Basics. <https://www.energy.gov/eere/solar/solar-radiation-basics>, (accessed: 19/10/2023).
2. K.A. Ibrahim, P.M. Gyuk, S. Aliyu. The Effect of Solar Irradiation on Solar Cells., Science World Journal, vol 14 (No.1), 2019.
3. Ramez Abdallah, Adel Juaidi , Tareq Salameh, Mejdi Jeguirim, Hüseyin Çamur, Youssef Kassem , Samer Abdala. Estimation of solar irradiation and optimum tilt angles for south-facing surfaces in the United Arab Emirates: a case study using PVGIS and PVWatts., Recent Advances in Renewable Energy Technologies, 2022.
4. A. Jakoplić, D. Franković. V. Kirinčić. T. Playšić. Benefits of short-term photovoltaic power production forecasting to the power system., Optimization and Engineering, 2020.
5. Earth Science Communications Team, NASA's Jet Propulsion Laboratory. The Causes of Climate Change. <https://climate.nasa.gov/causes/>, (accessed: 19/10/2023).
6. Espinar, B., Aznarte, J. L., Girard, R., Moussa, A. M., Kariniotakis, G. Photovoltaic Forecasting: A state of the art., In 5th European PV-Hybrid and Mini-Grid Conference (pp. Pages-250). OTTI-Ostbayerisches Technologie-Transfer-Institut. 2010.
7. Moreno-Munoz, A., De La Rosa, J. J. G., Posadillo, R., Bellido, F. Very short term forecasting of solar radiation., 33rd IEEE Photovoltaic Specialists Conference (pp. 1-5). IEEE. 2008.

8. Anderson, D., Leach, M. Harvesting and redistributing renewable energy: on the role of gas and electricity grids to overcome intermittency through the generation and storage of hydrogen., In *Energy Policy* (Vol. 32, Issue 14, pp. 1603-1614). Elsevier BV. 2004.
9. Hammer, A., Heinemann, D., Lorenz, E., Lückehe, B. Short-term forecasting of solar radiation: a statistical approach using satellite data., In *Solar Energy* (Vol. 67, Issues 1-3, pp. 139-150). Elsevier BV. 1999.
10. Frías-Paredes, L., Mallor, F., Gastón-Romeo, M., León, T. Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors., In *Energy Conversion and Management* (Vol. 142, pp. 533-546). Elsevier BV. 2017.
11. Bautista Carrascosa, Inmaculada. Cálculo de la radiación solar extraterrestre en función de la latitud y la declinación solar. <https://m.riunet.upv.es/bitstream/handle/10251/68296/Bautista%20-%20C%C3%A1lculo%20de%20la%20radiaci%C3%B3n%20solar%20extraterrestre%20en%20funci%C3%B3n%20de%20la%20latitud%20y%20la%20declinaci%C3%B3n...pdf?sequence=1&isAllowed=y>, (accessed: 19/10/2023).
12. Bamisile, O., Cai, D., Oluwasanmi, A., Ejiyi, C., Ukwuoma, C. C., Ojo, O., Mukhtar, M., Huang, Q. Comprehensive assessment, review, and comparison of AI models for solar irradiance prediction based on different time/estimation intervals., *Scientific Reports* (Vol. 12, Issue 1). Springer Science and Business Media LLC, 2022.
13. Neha Sehrawat, Sahil Vashisht, Amritpal Singh. Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison., *International Journal of Intelligent Networks*.