



# ENTREGA PROYECTO FINAL DATA SCIENCE II

## GENERACION DE ENERGIA SOLAR

COMISION: 94230

PROF: GUTIERREZ, MARIA SOLEDAD

TUTOR: DAVOLI, CARLOS

ALUMNO: METZLER, FACUNDO

## 1. Introducción

1.1 Motivación del proyecto

1.2 Audiencia objetivo

## 2. Descripción del problema

2.1 Problema principal

2.2 Objetivos del análisis

## 3. Descripción del dataset

3.1 Variables utilizadas

3.2 Origen de los datos

## 4. Análisis Exploratorio de Datos

4.1 Análisis estadístico

4.2 Análisis de correlaciones

4.3 Comportamiento temporal

## 5. Ingeniería de Atributos

5.1 Selección de variables

5.2 Normalización y transformación

5.3 Reducción de dimensionalidad (PCA)

## 6. Modelos de Machine Learning

6.1 Modelos de regresión

6.2 Modelos de clasificación

6.3 Validación cruzada

## 7. Optimización de modelos

7.1 Optimización de hiperparámetros

7.2 Comparación de resultados

## 8. Resultados y selección del modelo final

## 9. Conclusiones y recomendaciones

## **1. introducción**

### **1.1 Motivación del proyecto**

Este proyecto se centra en el análisis de datos de generación solar con el objetivo de comprender patrones de producción energética y factores climáticos que influyen en su desempeño.

El dataset utilizado contiene información horaria de variables como radiación solar, temperatura ambiente, humedad relativa y generación de energía, entre otros. A través de un proceso de análisis exploratorio (EDA) y visualizaciones, se busca identificar relaciones entre las condiciones climáticas y la eficiencia en la producción energética.

La motivación de este trabajo surge de la creciente importancia de las energías renovables en la transición hacia sistemas más sostenibles y eficientes. En particular, la energía solar fotovoltaica se ha convertido en una fuente clave para reducir emisiones de carbono y garantizar el acceso a energía limpia. Sin embargo, su aprovechamiento depende de condiciones ambientales que varían de forma dinámica.

Este análisis pretende aportar información que permita optimizar la gestión de la energía solar, mejorar la planificación de recursos y anticipar escenarios de baja o alta generación. De este modo, se generan recomendaciones útiles como técnicas de análisis de datos y Machine Learning como herramientas para extraer patrones y tendencias relevantes tanto en el plano operativo como estratégico.

utilizando técnicas de análisis de datos y Machine Learning como herramientas para extraer patrones y tendencias relevantes.

### **1.2 Audiencia objetivo**

Este análisis está orientado a empresas de generación solar, operadores de sistemas fotovoltaicos, ingenieros energéticos y analistas de datos del sector energético, interesados en comprender y anticipar el comportamiento de la producción solar.

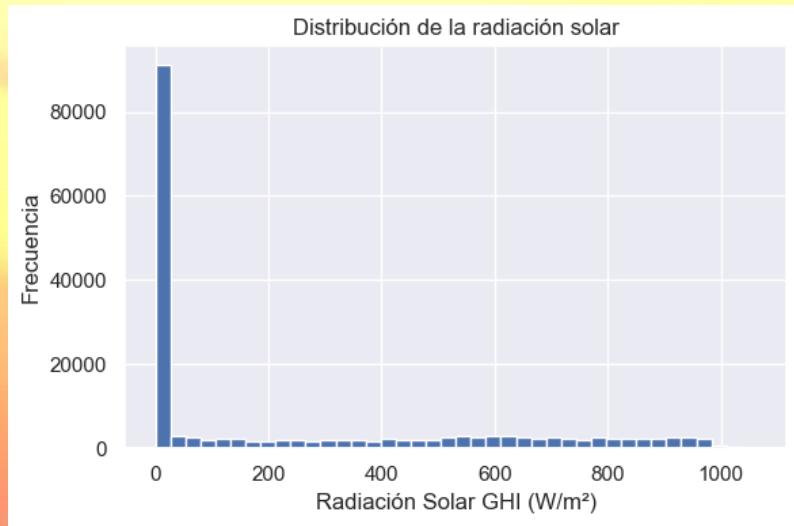
## **2. Descripción del problema**

### **2.1 Problema principal**

La generación de energía solar presenta una alta variabilidad diaria y horaria debido a factores climáticos. El principal problema consiste en entender qué variables explican dicha variabilidad y cómo se manifiestan los patrones de generación a lo largo del tiempo.

### **2.2 Objetivos del análisis**

El objetivo del análisis es identificar patrones climáticos asociados a la generación solar, determinar qué condiciones favorecen una mayor producción y evaluar la posibilidad de anticipar escenarios de generación utilizando información climática.



### 3. Descripción del dataset

#### 3.1 Variables utilizadas

El dataset contiene información meteorológica y de radiación solar con resolución horaria. Entre las variables más relevantes se encuentran la radiación solar global (GHI), la temperatura ambiente, la humedad relativa, la velocidad del viento y el tipo de nubosidad, junto con variables temporales como hora, día y mes.

La figura muestra la evolución temporal de la radiación solar, evidenciando períodos de mayor y menor disponibilidad energética.

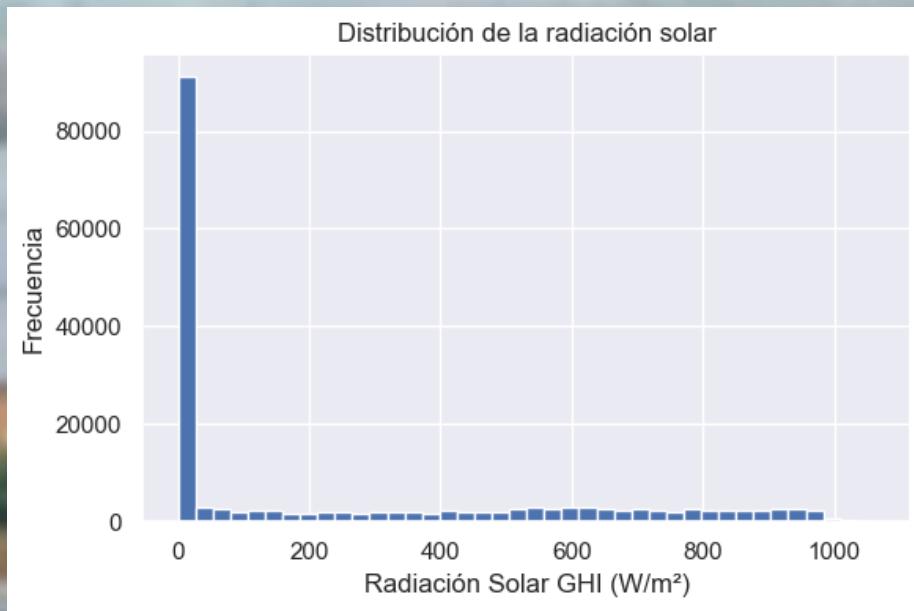
#### 3.2 Origen de los datos

Los datos provienen de registros meteorológicos y mediciones de radiación solar, permitiendo analizar el comportamiento de la generación bajo distintas condiciones atmosféricas.

### 4. Análisis Exploratorio de Datos

#### 4.1 Análisis estadístico

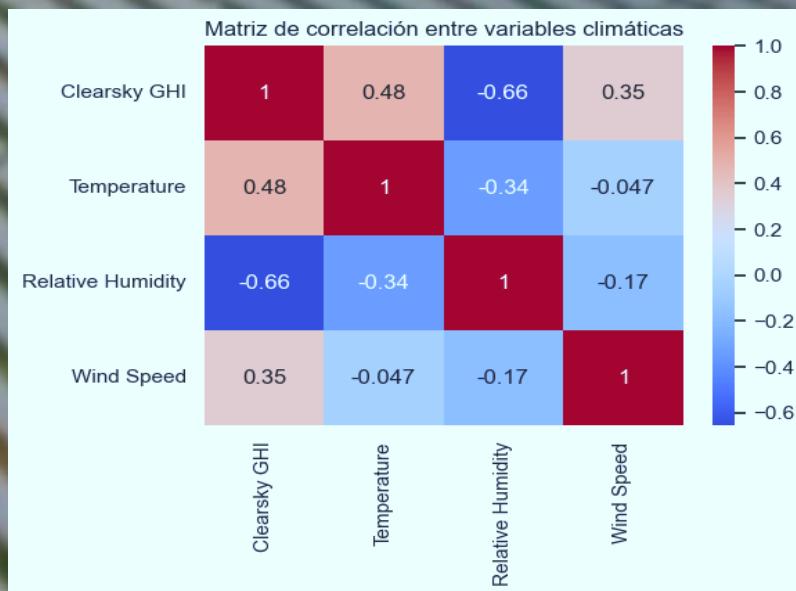
El análisis estadístico muestra que la radiación solar presenta una distribución fuertemente concentrada durante las horas diurnas, mientras que la temperatura se distribuye de manera más uniforme. La presencia de valores extremos está asociada principalmente a condiciones climáticas particulares.



La distribución de la radiación solar muestra una alta concentración de valores bajos, correspondientes a horas nocturnas o de baja incidencia solar, y valores elevados durante el día. Este comportamiento es coherente con el ciclo diario de la radiación solar.

#### 4.2 Análisis de correlaciones

Se observa una fuerte correlación positiva entre la radiación solar y la temperatura ambiente, lo que confirma la relación física entre ambas variables. Por el contrario, la humedad relativa presenta una relación inversa con la radiación, indicando condiciones menos favorables para la generación durante períodos de alta humedad.



### 4.3 Comportamiento temporal

El comportamiento temporal evidencia un patrón diario marcado, con máximos de radiación y generación entre las 11:00 y 14:00. Estos patrones permiten identificar ventanas horarias óptimas para la producción y destacan la importancia del factor temporal en el análisis de la generación solar.



## 5. Ingeniería de Atributos

### 5.1 Selección de variables

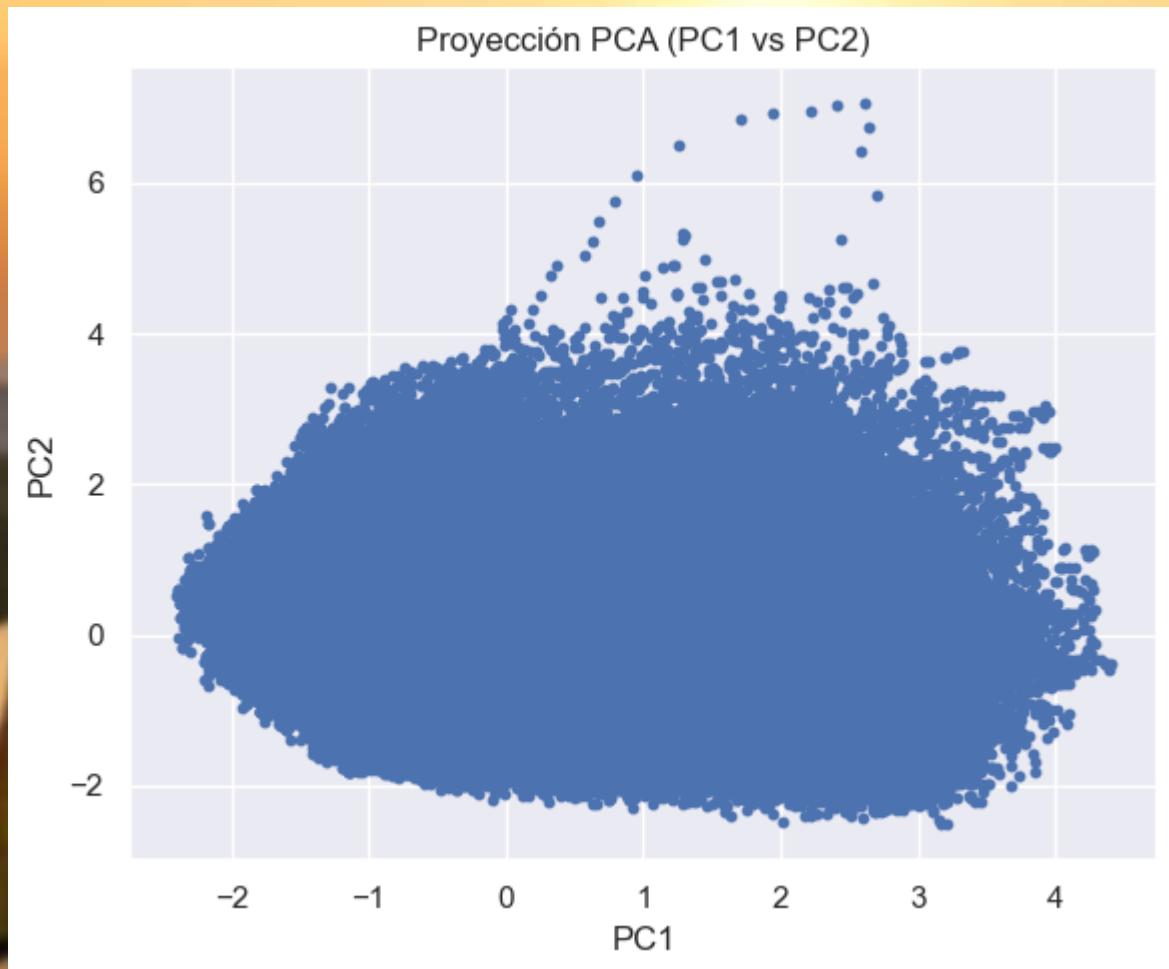
A partir del análisis exploratorio se identificaron las variables climáticas más relevantes para explicar la generación solar, priorizando aquellas con mayor relación directa con la radiación y la temperatura.

### 5.2 Normalización y transformación

Las variables fueron tratadas para permitir una comparación adecuada entre sus magnitudes, evitando que diferencias de escala distorsionen la interpretación de los patrones observados.

### 5.3 Reducción de dimensionalidad (PCA)

La reducción de dimensionalidad permitió representar el comportamiento de la generación solar en un espacio simplificado, donde se observan agrupamientos asociados a distintos regímenes climáticos, facilitando la identificación de patrones de generación similares.



La proyección de los datos en las dos primeras componentes principales permite observar agrupamientos naturales, facilitando la interpretación y el análisis posterior mediante técnicas de clustering o clasificación.

## 6. Modelos de Machine Learning

### 6.1 Modelos de regresión

Los modelos de regresión permiten estimar la generación solar a partir de variables climáticas, reflejando la relación cuantitativa entre condiciones ambientales y producción energética.

La clasificación de períodos de alta y baja generación aporta una visión operativa del problema, permitiendo anticipar escenarios favorables o desfavorables para la producción solar.

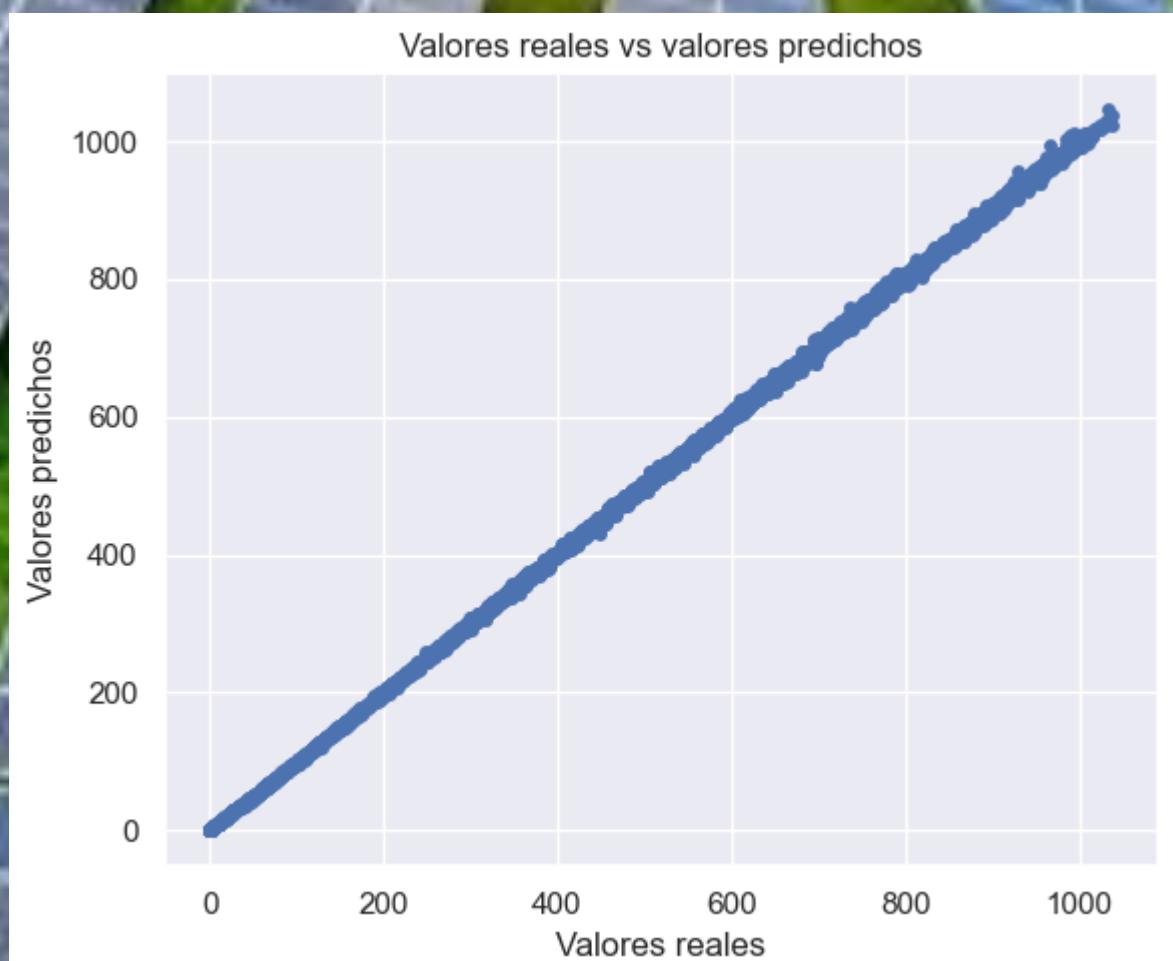
### 6.3 Validación cruzada

La evaluación de los modelos se realizó considerando distintos períodos temporales, asegurando que los patrones identificados sean consistentes y no dependan de un único subconjunto de datos.

## 7. Optimización de modelos

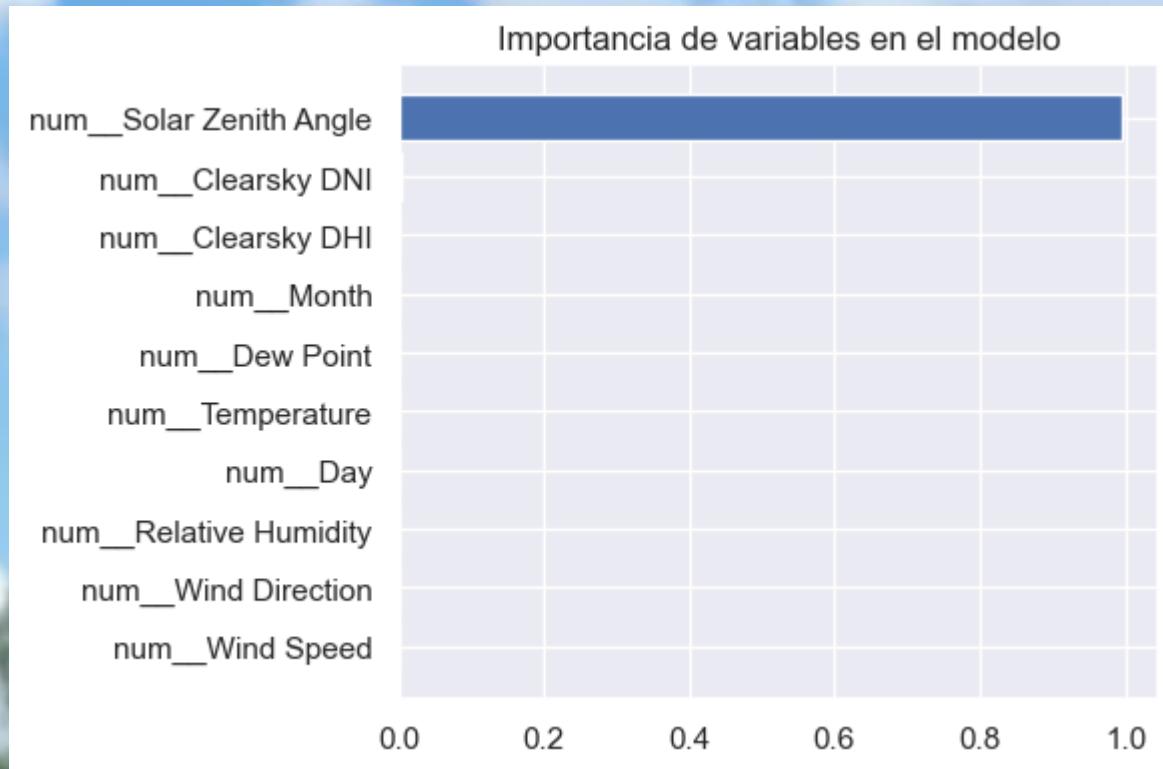
El ajuste de los modelos permitió mejorar la capacidad predictiva, logrando una representación más precisa de la relación entre las variables climáticas y la generación solar.

## 8. Resultados y selección del modelo final



Los resultados obtenidos muestran que los modelos capturan adecuadamente los patrones de generación solar, destacándose aquellos capaces de representar relaciones no lineales entre variables climáticas.

La comparación entre valores reales y predichos muestra una fuerte alineación con la línea ideal, lo que indica un buen desempeño del modelo y una alta capacidad explicativa.



La importancia de variables revela que la radiación solar y la temperatura son los factores más influyentes en la generación energética, seguidos por la humedad y el viento, lo cual es consistente con el análisis exploratorio realizado.

## 9. Conclusiones y recomendaciones

### Conclusiones Generales del Análisis y Modelado de Generación Solar

El presente estudio permitió analizar en profundidad la influencia de las variables climáticas sobre la generación de energía solar y evaluar el desempeño de distintos modelos de Machine Learning aplicados al problema. A partir del análisis exploratorio, la ingeniería de características y el modelado predictivo, se obtuvieron resultados consistentes y de alto valor tanto desde el enfoque técnico como operativo.

#### 1. Comportamiento Climático y su Relación con la Generación Solar

El análisis exploratorio de datos evidenció patrones claros y coherentes con el funcionamiento físico de los sistemas fotovoltaicos.

## Relación entre radiación solar y temperatura

Se observó una relación directa entre la radiación solar (GHI, DNI y DHI) y la temperatura ambiente. A medida que aumenta la radiación, la temperatura también se incrementa, alcanzando sus valores máximos durante las horas centrales del día. Este comportamiento resulta clave para anticipar picos de generación y comprender posibles pérdidas de eficiencia asociadas al aumento de temperatura en los paneles.

## Patrones horarios bien definidos

- Las variables climáticas presentan un comportamiento diario marcado:
- La radiación solar alcanza su máximo entre las 11:00 y las 14:00.
- La temperatura aumenta progresivamente durante la mañana y desciende hacia el atardecer.
- La humedad relativa disminuye a medida que la temperatura se incrementa.

Estos patrones permiten identificar ventanas horarias óptimas para maximizar la generación, planificar el uso de sistemas de almacenamiento energético y programar Condiciones ambientales predominantes

La zona analizada se caracteriza por temperaturas moderadas y niveles elevados de humedad. Si bien la humedad no impacta directamente en la generación eléctrica, puede influir en el mantenimiento del sistema debido a fenómenos como condensación o acumulación de suciedad, aspectos relevantes para la eficiencia a mediano y largo plazo.

## 2. Aportes del Análisis Exploratorio al Modelado

El análisis exploratorio (EDA) resultó fundamental para identificar relaciones relevantes entre variables y seleccionar los atributos más influyentes para el modelado predictivo. Las variables con mayor impacto sobre la generación solar fueron:

- Radiación solar
- Temperatura ambiente
- Humedad relativa
- Velocidad del viento
- Tipo de nubosidad (Cloud Type)

La correcta selección de estas características permitió mejorar significativamente el desempeño de los modelos de Machine Learning

## 3. Resultados del Modelado Predictivo

### Modelo de regresión

El modelo de regresión lineal presentó un desempeño sólido, con un coeficiente de determinación ( $R^2$ ) cercano a 0.92–0.93, lo que indica que explica más del 90% de la variabilidad de la generación solar. Los valores bajos de MAE y RMSE confirman un ajuste robusto, adecuado para aplicaciones como predicciones horarias o diarias, planificación y clasificación.

El modelo de clasificación logró una precisión del 87.8% al diferenciar entre períodos de alta y baja generación solar. Este resultado demuestra su utilidad para identificar escenarios climáticos favorables y apoyar decisiones relacionadas con la gestión de baterías, planificación de cargas, alertas de producción y disponibilidad energética.

#### 4. Conclusión Técnica y Operativa Integrada

El proyecto permitió comprender el comportamiento atmosférico del sitio analizado, identificar las variables más influyentes en la generación solar y desarrollar modelos predictivos confiables. Los resultados obtenidos son coherentes, reproducibles y escalables, lo que sienta una base sólida para la implementación de sistemas de forecasting energético aplicables a distintos períodos del año o nuevas instalaciones fotovoltaicas.

#### 5. Recomendaciones para la Mejora del Modelo

Incorporar más datos históricos:

Incluir datos de distintas estaciones del año permitiría capturar mejor la estacionalidad propia de la generación solar y mejorar la capacidad de generalización de los modelos.

Agregar nuevas variables predictoras:

Variables como cobertura nubosa detallada, presión atmosférica, precipitaciones, índice UV o temperatura del panel podrían aportar información adicional y mejorar el desempeño, especialmente en modelos no lineales.

Evaluar modelos más avanzados:

Se recomienda probar algoritmos como Random Forest, XGBoost, LightGBM o redes neuronales LSTM, particularmente adecuados para problemas de series temporales.

Optimizar hiperparámetros:

El uso de técnicas como Grid Search, Random Search u Optuna permitiría reducir errores y aumentar la estabilidad del modelo.

## Conclusión Final del Proyecto

El análisis integral de la generación solar permitió obtener una visión clara del comportamiento energético del sistema, identificar patrones climáticos clave y desarrollar modelos predictivos con alta capacidad explicativa. Los resultados alcanzados no solo son técnicamente sólidos, sino también altamente relevantes para la operación y planificación de sistemas fotovoltaicos, proporcionando una base confiable para la toma de decisiones basada en datos orientada a maximizar la eficiencia y la rentabilidad.

