



Regresión Lineal Múltiple para Predicción de Energía Solar

Un Enfoque de Machine Learning Aplicado a la
Ingeniería

Elizabeth Hosten, Camilo Calderin

Universidad Pontificia Bolivariana - Montería

Introducción

Objetivos del Proyecto

1

Modelado Predictivo

Desarrollar un modelo de **Regresión Lineal Múltiple** para estimar la producción de energía FV.

2

Comparación Geográfica

Contrastar el rendimiento del modelo entre dos ubicaciones climáticas distintas: **Dubai y Montería**.

3

Evaluación del Modelo

Medir la precisión y el rendimiento utilizando métricas clave como **R², MAE y RMSE**.

4

Análisis de Variables

Identificar y cuantificar la influencia de las variables ambientales en cada contexto geográfico.

Metodología: Del Dato Crudo a la Predicción



1. Preparación de Datos

Carga de archivos CSV (datos de 2022 para Dubai y Montería). Ajuste de zonas horarias y cálculo de la producción real utilizando un **Modelo Físico (POO)** para el panel JinkoSolar Tiger Pro 540W.



3. Evaluación del Modelo

Medición del rendimiento con métricas estadísticas: **R²** (variabilidad explicada), **MAE** (error promedio), y **RMSE** (penalización de errores grandes).

2. Construcción del Modelo

División de datos: **80% para entrenamiento** y 20% para prueba. Entrenamiento del modelo RLM usando las 6 variables ambientales como input.

4. Análisis e Interpretación

Identificación de las variables más influyentes. Comparación detallada de los resultados entre las dos ubicaciones y generación de visualizaciones.

Resultados de la Modelización: Dubai (Clima Desértico)

El modelo de Regresión Lineal Múltiple mostró un ajuste **EXCELENTE** en el entorno de alta radiación constante de Dubai.

0.9995

Coeficiente R²

99.95% de la variabilidad de la producción es explicada por el modelo.

2.72 W

Error Absoluto Medio (MAE)

Error promedio muy bajo, equivalente a solo el 0.50% de la potencia nominal (540W).

Conclusión Clave

GHI (Radiación Horizontal Global) es la variable dominante, representando el 43.8% de la importancia total. La radiación solar es el factor principal en este clima.



Resultados de la Modelización: Montería (Clima Tropical)

Sorprendentemente, el modelo mostró un ajuste ligeramente superior en Montería, aunque el factor dominante cambió drásticamente debido a la naturaleza del clima tropical.



99.97%

**Variabilidad
Explicada**

R^2 de 0.9997, indicando un ajuste aún mejor que Dubai.

1.70 W

**Error Promedio
Mínimo**

El MAE más bajo, representando solo el 0.31% de la potencia nominal.

Conclusión Clave

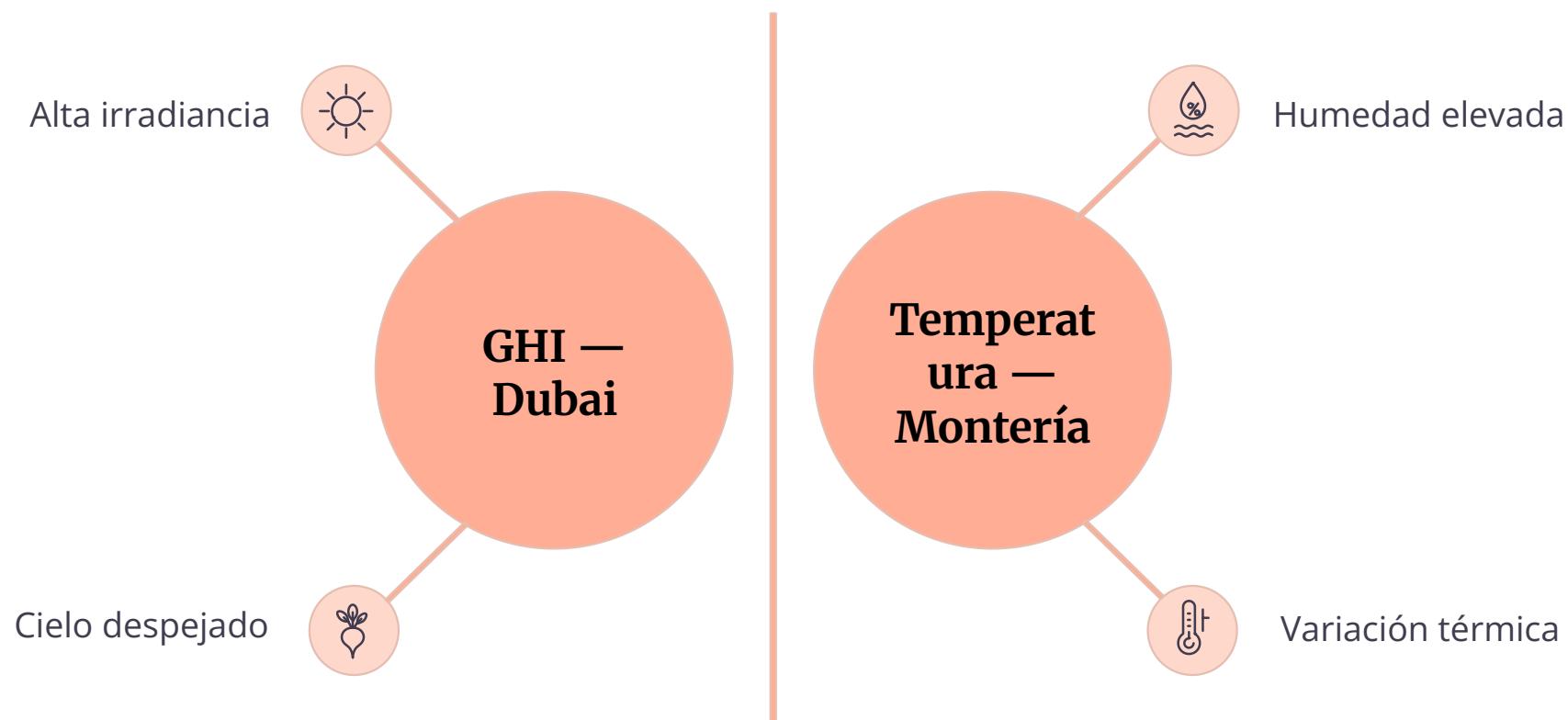
En Montería, la **Temperatura** es la variable más influyente (45.3%). El clima tropical, con su alta temperatura y humedad, impacta la eficiencia del panel más que la radiación directa.

Comparación Detallada: Dubai vs. Montería

La comparación de métricas revela que el modelo predictivo funciona excepcionalmente bien en ambos contextos, pero con una ligera ventaja para Montería.

Métrica	Dubai	Montería	Análisis
R ² (Ajuste)	0.9995	0.9997	Montería logra explicar el 0.02% adicional de variabilidad.
MAE (Error Absoluto)	2.72 W	1.70 W	Montería tiene un 37.5% menor error promedio.
Variable Principal	GHI (43.8%)	Temperatura (45.3%)	Factores climáticos influyentes son distintos en cada región.

Impacto de las Variables Dominantes



Interpretación y Significado de los Coeficientes

Los resultados confirman que la Regresión Lineal Múltiple fue capaz de "aprender" la física subyacente del comportamiento del panel solar.

1

Altísima Relevancia (R^2)

El R^2 superior a 0.999 indica que **casi toda la variabilidad** de la producción está contenida en las 6 variables seleccionadas. El modelo es un sustituto estadístico excepcional del modelo físico.

2

Significado de los Coeficientes (β)

Los coeficientes de regresión revelan las relaciones causales:

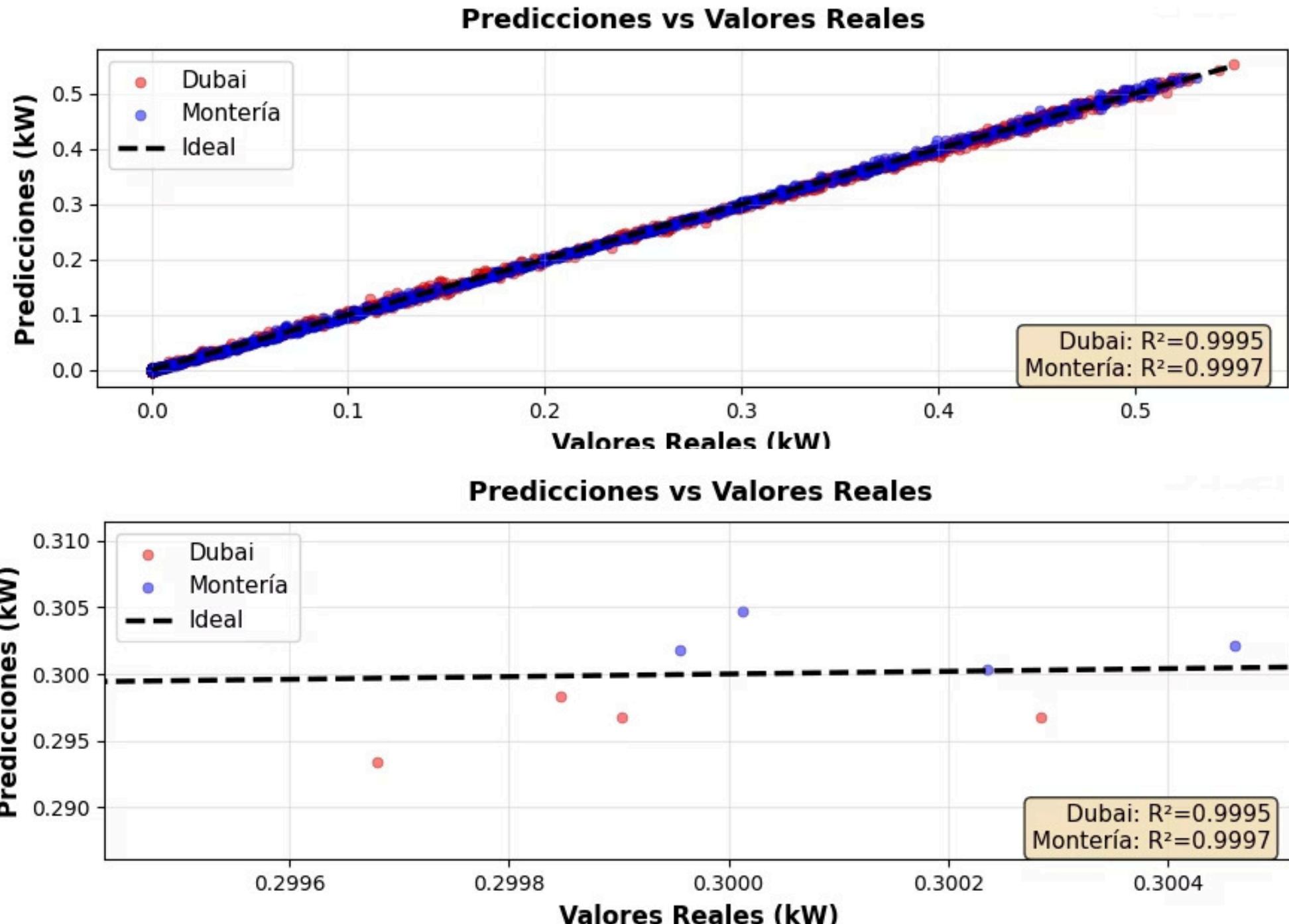
- Valores **Positivos** (ej. GHI): Aumentan la producción.
- Valores **Negativos** (ej. Temperatura): Reducen la producción por ineficiencia.

3

Margen de Error

Con un error promedio (MAE) de solo 1.70 W a 2.72 W en un panel de 540 W, la predicción es altamente precisa para aplicaciones de pronóstico energético.

Visualizaciones y Código



Estructura del Proyecto en Python

</>

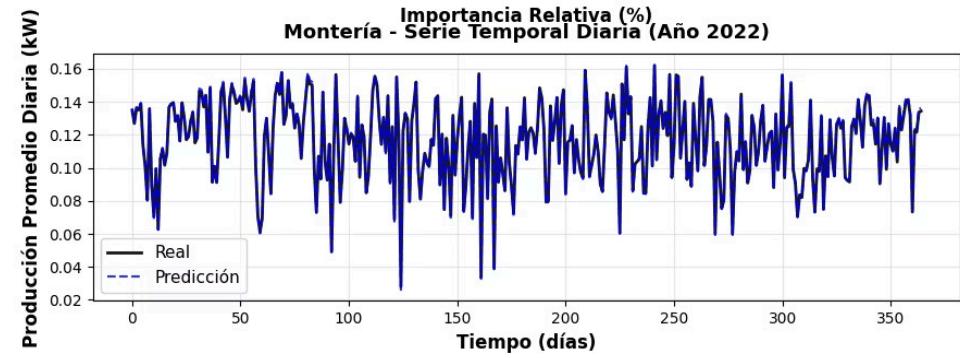
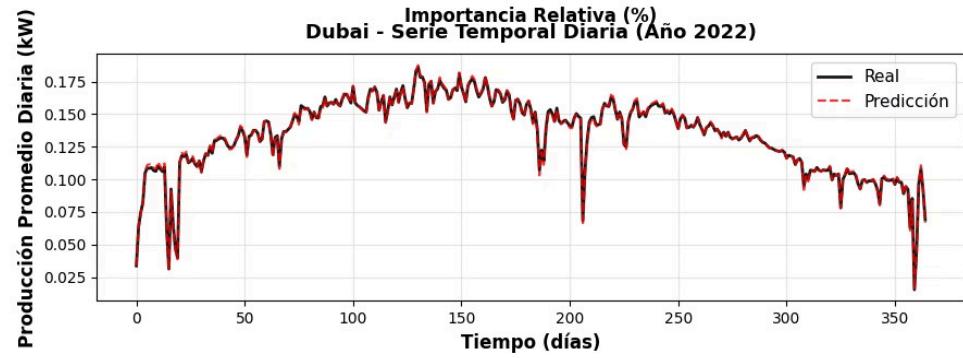
Modelo Físico (Clase POO)

Clase `PanelSolarFotovoltaico` implementando la física de producción, eficiencia y el coeficiente de temperatura.



Modelo Predictivo (scikit-learn)

Implementación de `LinearRegression()` y la metodología `train_test_split` para el entrenamiento y prueba del modelo ML.



Conclusiones y Futuro del Proyecto

Validación de ML Supervisado

La Regresión Lineal Múltiple es **altamente adecuada** para predecir la producción de energía FV en diferentes climas, logrando precisiones superiores al 99.9%.

Ruta a la Optimización

Explorar modelos más complejos (Random Forest, XGBoost) y añadir variables predictoras (viento, presión) para mejorar la robustez y capacidad de generalización.

Aplicaciones Prácticas

Los modelos desarrollados son la base para sistemas de **pronóstico de energía**, optimización de la gestión de la red y detección temprana de anomalías en paneles solares.