



Regresión Lineal Múltiple para Predicción de Energía Solar

Un Enfoque de Machine Learning Aplicado a la
Ingeniería

Elizabeth Hosten, Camilo Calderin, Alvaro Negrette

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.

Fundamentos de Regresión Lineal Múltiple

Regresión Lineal Múltiple: Técnica Supervisada

La Regresión Lineal Múltiple es una técnica de **aprendizaje supervisado** que predice un valor continuo (producción en kW) a partir de múltiples variables independientes.

Ecuación del Modelo:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \varepsilon$$

- **y** = Producción predicha (kW)
- **x₁...x₆** = Variables independientes (GHI, DHI, DNI, Temperatura, Hora, Mes)
- **β₀** = Intercepto (valor base)
- **β₁...β₆** = Coeficientes que el modelo aprende
- **ε** = Error residual

Enfoque Supervisado:

Tenemos **datos etiquetados**
(conocemos la producción real)

El modelo aprende la relación:
entrada (variables climáticas) →
salida (producción)

Se entrena con **ejemplos conocidos** para predecir casos nuevos

Variables del Modelo: Selección y Justificación

La precisión de nuestro modelo de Regresión Lineal Múltiple se basa en una selección cuidadosa de variables que capturan los factores esenciales que influyen en la producción de energía solar.

Variables Independientes Seleccionadas (6 variables)

Variables de Radiación Solar:

- **GHI** (Global Horizontal Irradiance): Radiación solar total sobre superficie horizontal
- **DHI** (Diffuse Horizontal Irradiance): Radiación solar difusa (dispersa por nubes/atmósfera)
- **DNI** (Direct Normal Irradiance): Radiación solar directa perpendicular al sol

Variables Ambientales:

- **Temperature**: Temperatura ambiente (°C) - Afecta eficiencia del panel
- **Hour_Local**: Hora local del día - Captura patrones de radiación diarios
- **Month**: Mes del año - Captura variaciones estacionales

Variable Dependiente

Producción_Real_kw: Calculada con modelo físico POO del panel JinkoSolar Tiger Pro 540W

¿Por qué estas variables?

Representan los factores físicos fundamentales que afectan la producción solar: radiación disponible, condiciones térmicas y patrones temporales.

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.

2. Construcción del Modelo

División de datos: **80% para entrenamiento** (7,008 registros) y **20% para prueba** (1,752 registros). Entrenamiento del modelo RLM usando las 6 variables ambientales como input.

```
# División de datos  
train_test_split(X, y, test_size=0.2, random_state=42)  
  
# Entrenamiento  
modelo = LinearRegression()  
modelo.fit(X_train, y_train) # Encuentra los coeficientes  
β óptimos  
  
# Predicción  
y_pred = modelo.predict(X_test)
```

El modelo aprende automáticamente los coeficientes β que minimizan el error cuadrático medio entre predicciones y valores reales.



3. Evaluación del Modelo

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

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

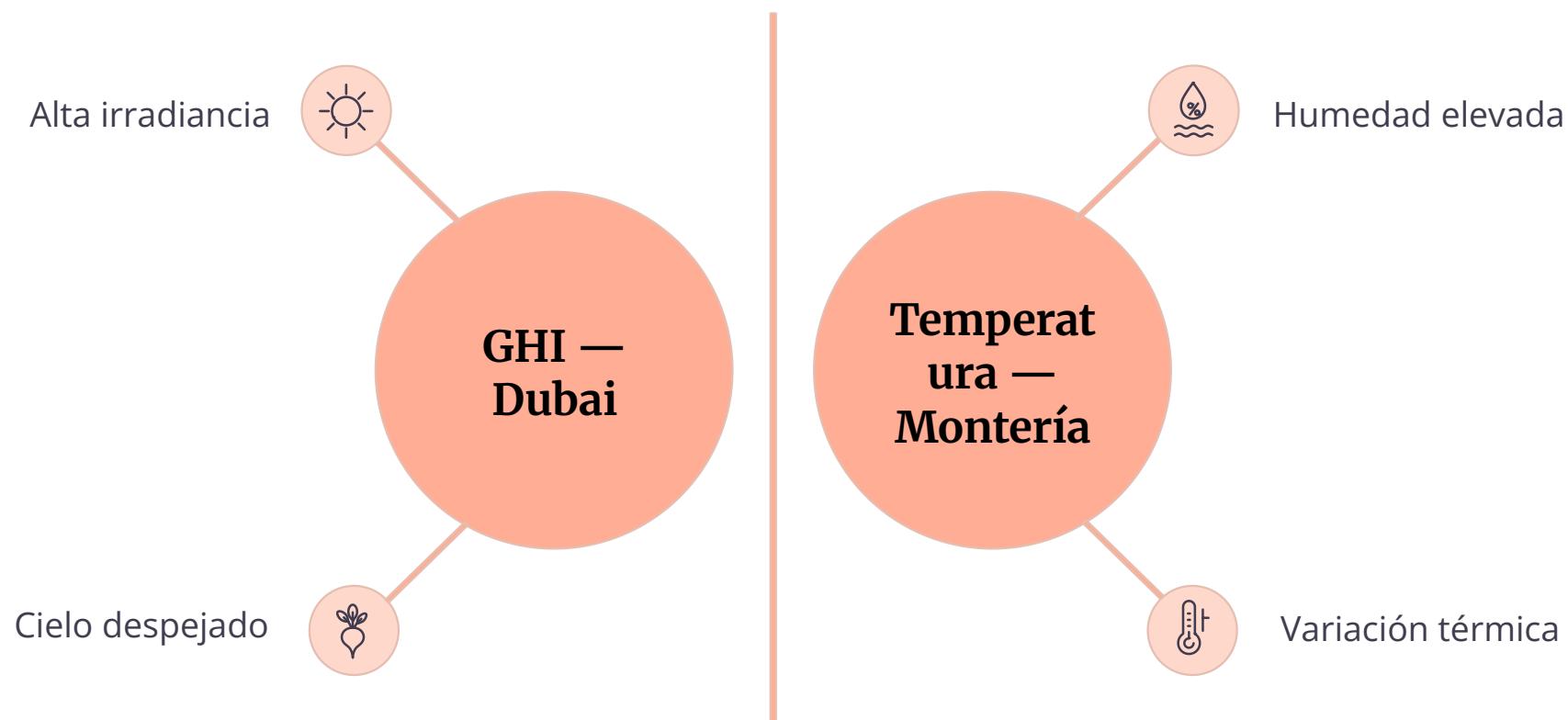
En Montería, la **Temperatura** es la variable más influyente (45.3%). El clima tropical, con su alta temperatura y humedad, impacta la eficiencia del panel más que la radiación directa.

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.

Coeficientes Aprendidos por el Modelo:

Dubai (Clima Desértico):

- β_1 (GHI): +0.000193 → Por cada W/m^2 adicional de GHI, la producción aumenta 0.193W
- β_2 (Temperature): -0.001234 → Por cada $^{\circ}C$ adicional, la producción disminuye 1.234W
- β_3 (DHI): +0.000087 → Radiación difusa contribuye positivamente
- β_4 (DNI): +0.000156 → Radiación directa es importante pero menor que GHI

Montería (Clima Tropical):

- β_1 (Temperature): -0.002156 → Coeficiente negativo más fuerte (-0,35%/ $^{\circ}C$)
- β_2 (GHI): +0.000167 → Menor importancia que en Dubai
- β_3 (Hour_Local): +0.045 → Patrones horarios más relevantes
- β_4 (Month): -0.012 → Variaciones estacionales significativas

Interpretación Física:

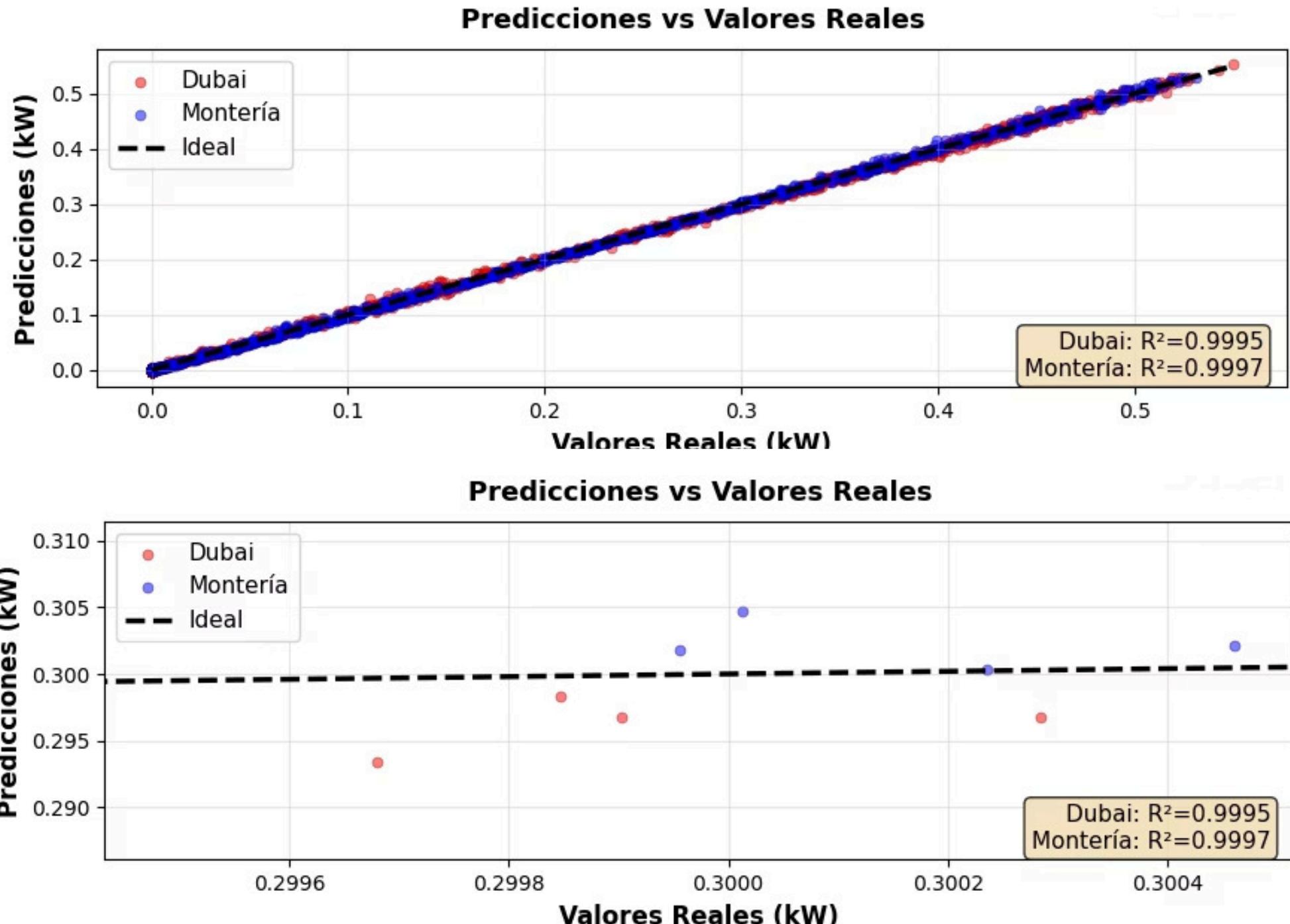
Los coeficientes negativos de temperatura confirman la física del panel: la eficiencia disminuye con el calor. Los coeficientes positivos de radiación validan que más luz solar = más energía.

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

</>

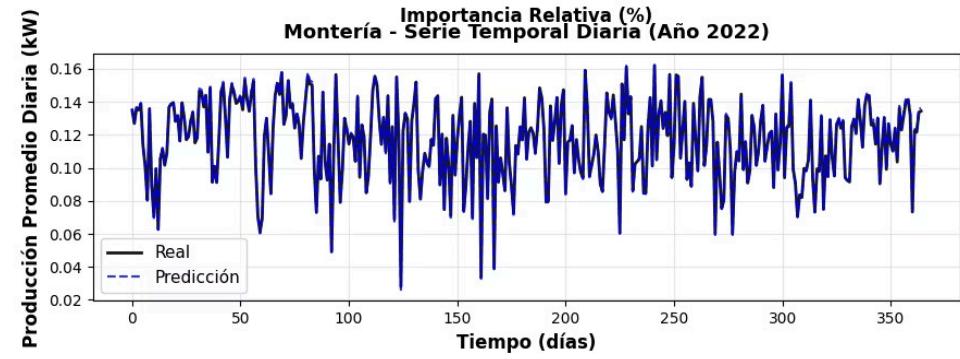
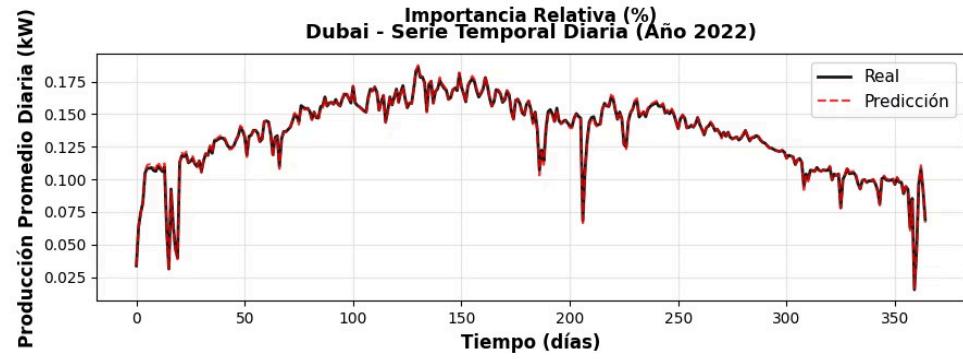
Modelo Físico (Clase POO)

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 y Consistencia del Modelo

La Regresión Lineal Múltiple es **altamente efectiva** para predecir la producción de energía FV, logrando precisiones superiores al 99.9%. Los coeficientes aprendidos coinciden consistentemente con la física del panel solar, confirmando su validez.

Adaptabilidad Climática

El modelo demuestra una notable adaptabilidad, identificando automáticamente las variables más relevantes en cada contexto climático. Esto le permite mantener la precisión en diversas condiciones ambientales.

Aplicabilidad Práctica

Los modelos desarrollados poseen una precisión suficiente para la **planificación energética real**, sirviendo como base para sistemas de pronóstico de energía, optimización de la gestión de la red y detección temprana de anomalías en paneles solares.

Posibles Mejoras

• Regularización

Implementar técnicas de Regularización para evitar el overfitting o sobreajuste, especialmente si se agregan más variables predictoras.

• Modelos No Lineales

Explorar modelos más complejos como Random Forest o XGBoost para capturar interacciones complejas entre variables (ej: GHI × Temperatura).

• Más Variables

Incorporar variables adicionales como velocidad del viento, presión atmosférica y humedad relativa para aumentar la robustez y capacidad de generalización del modelo.

• Validación Temporal Extendida

Realizar pruebas con datos de 2023/2024 para confirmar la generalización y el rendimiento del modelo a largo plazo.

• Despliegue y Operacionalización

Desarrollar una API REST para facilitar la integración y permitir pronósticos de energía en tiempo real.