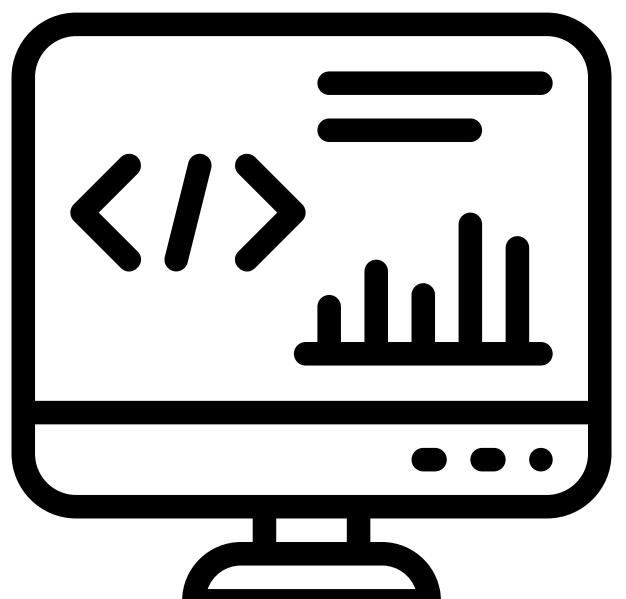


EVALUACION #5

REGRESION LINEAL



Equipo:
Samuel González
Alejandro González
Damaso Doria

INTRODUCCIÓN

- Pandas → para el manejo y limpieza de datos.
- NumPy → para operaciones matemáticas y vectorización.
- Matplotlib y Seaborn → para visualización y análisis gráfico.
- Scikit-learn → para implementar el modelo de regresión lineal múltiple y calcular métricas de desempeño.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
from sklearn.preprocessing import StandardScaler
```

OBJETIVOS

- Aplicar regresión lineal múltiple con datos reales
- Crear modelos predictivos para ambas ubicaciones
- Comparar y evaluar el desempeño de los modelos
- Identificar variables más influyentes
- Realizar predicciones validadas

PROCESAMIENTO Y LIMPIEZA DE DATOS

- Se cargan los conjuntos de datos en estructuras de pandas para su manipulación.
- Se revisan valores nulos, tipos de datos y se seleccionan columnas relevantes.
- Se preparan los datos para análisis estadístico y regresión lineal múltiple.

VISUALIZACIÓN GRÁFICAS

- Se generan diferentes tipos de gráficos con matplotlib:
- Predicciones vs valores reales en ambas ciudades
- Análisis de residuos
- Importancia de variables (coeficientes del modelo)

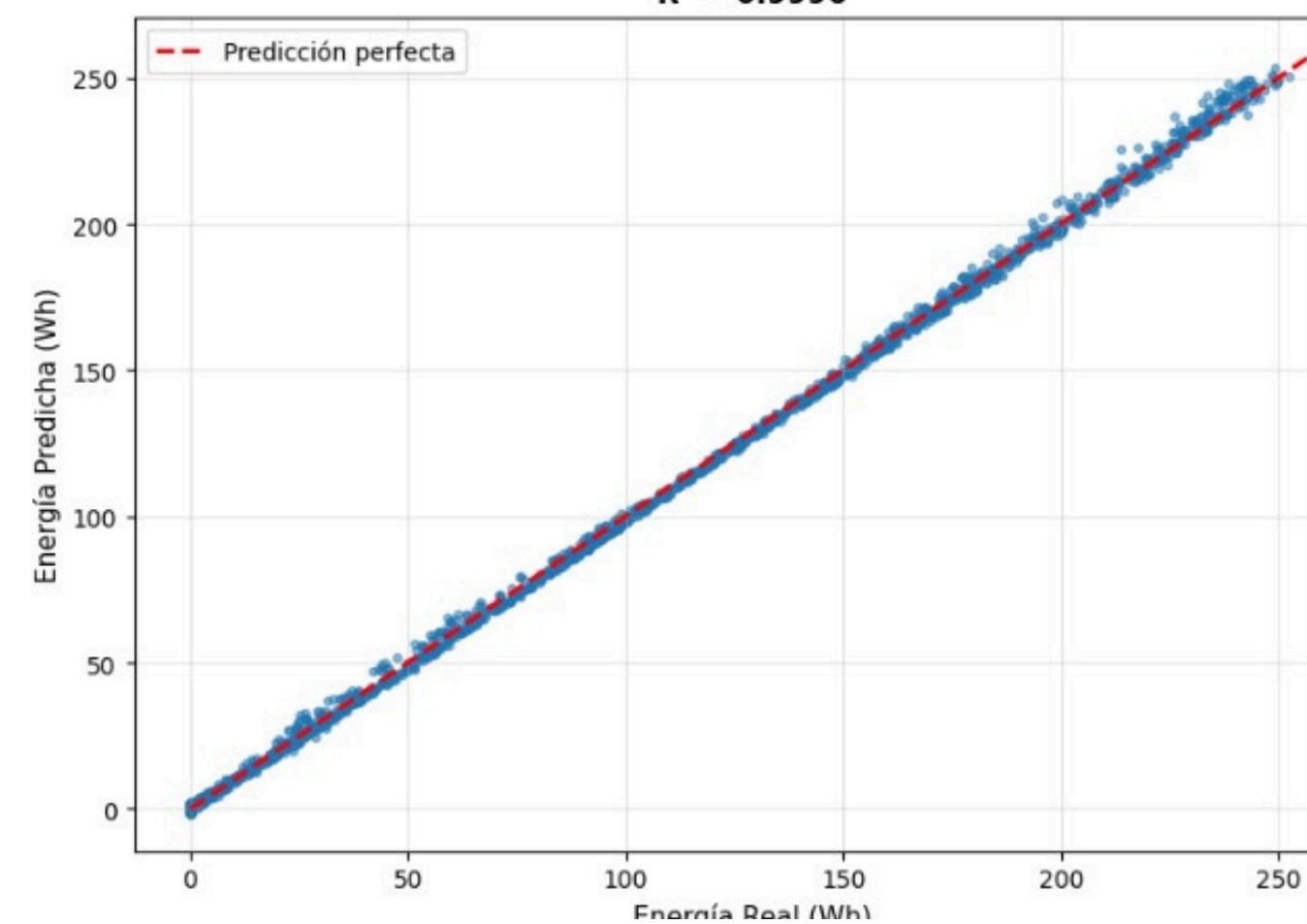
SIMULACIÓN CON PANEL SOLAR

- Este proyecto realiza un análisis comparativo de datos de radiación solar en dos ubicaciones distintas (Montería y Sídney), utilizando técnicas estadísticas y gráficas para comprender el comportamiento de la energía solar y su relación con diferentes variables meteorológicas.

Evaluación del Modelo de Regresión Lineal Múltiple

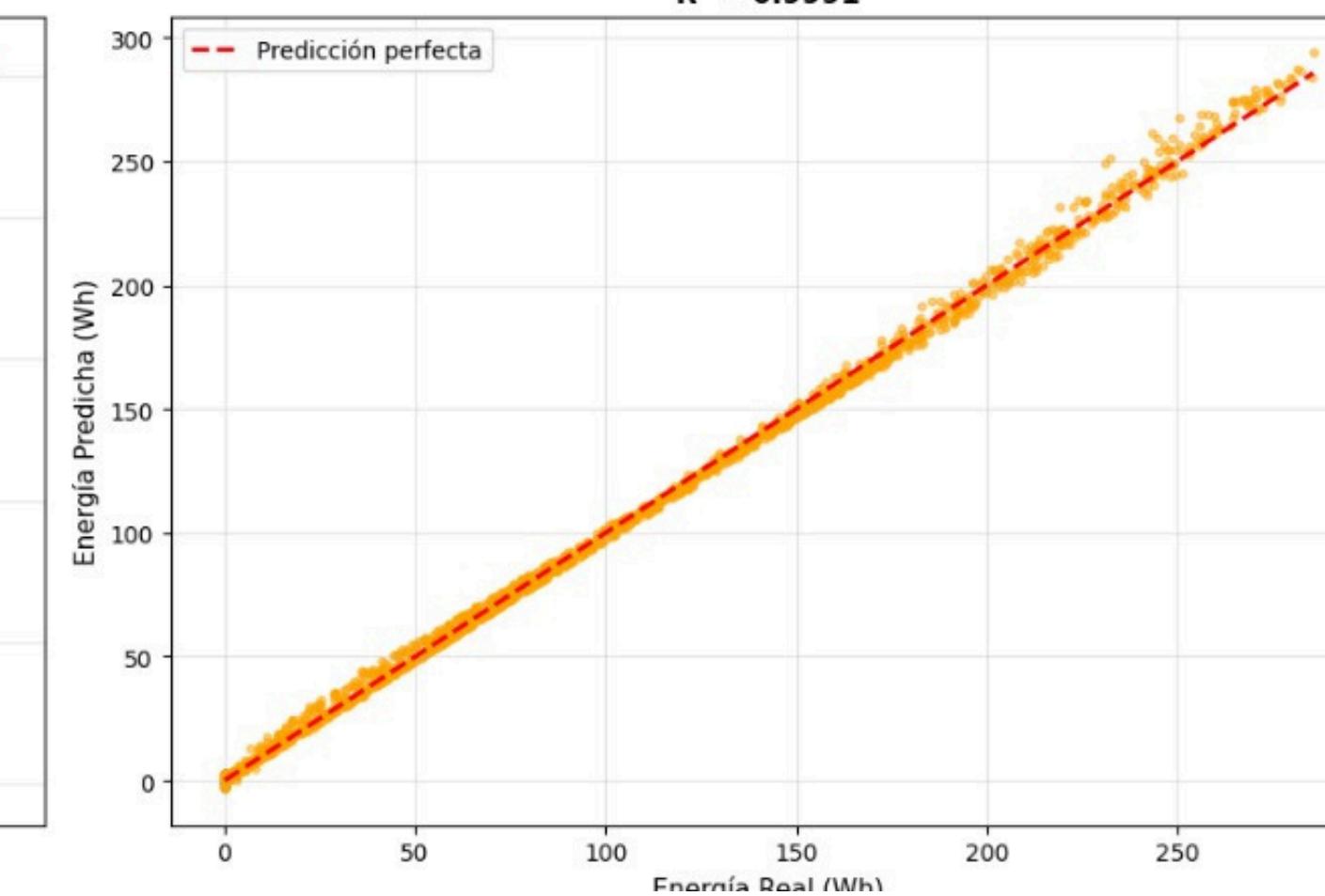
Montería - Predicciones vs Valores Reales

$R^2 = 0.9996$

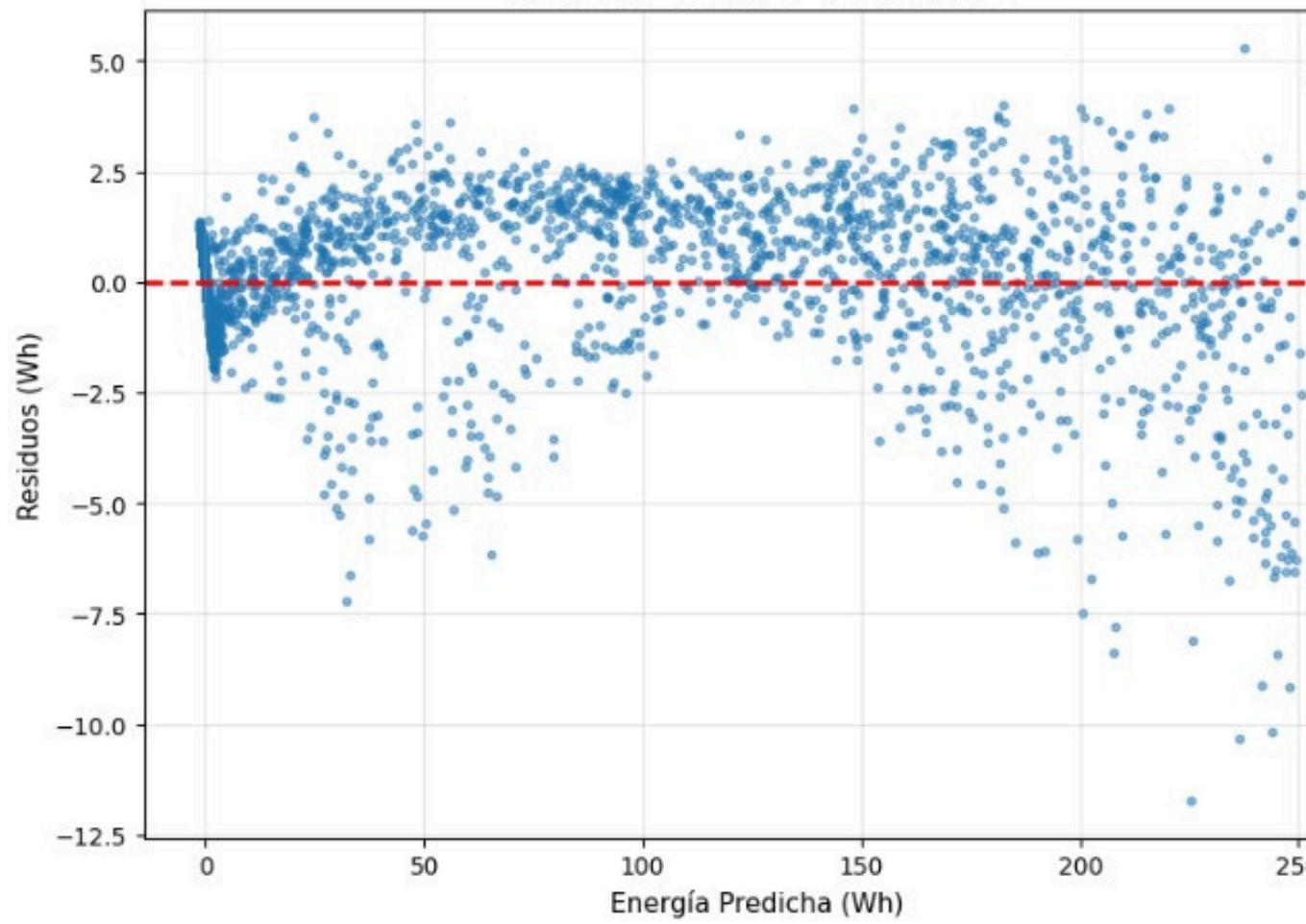


Sidney - Predicciones vs Valores Reales

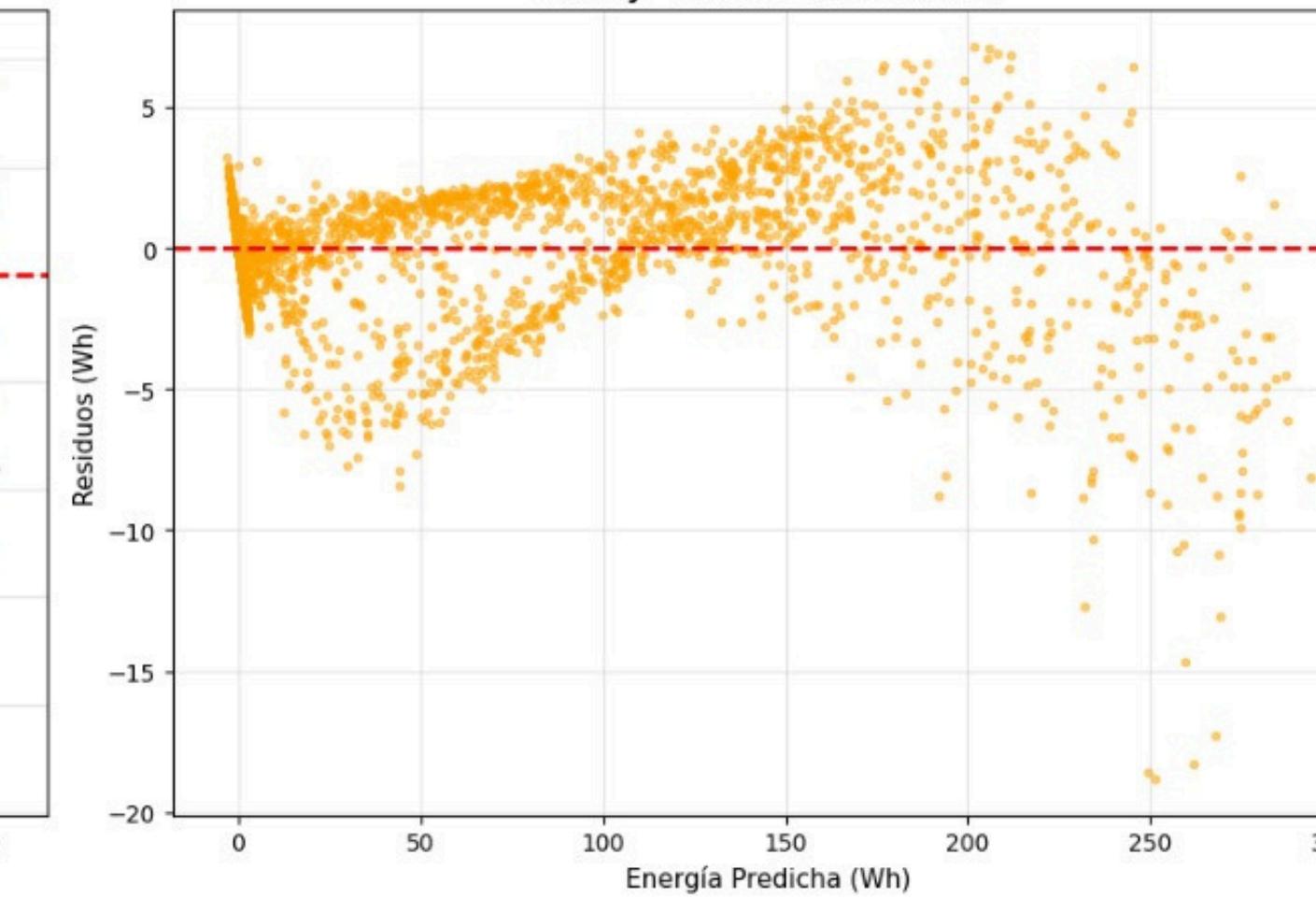
$R^2 = 0.9991$



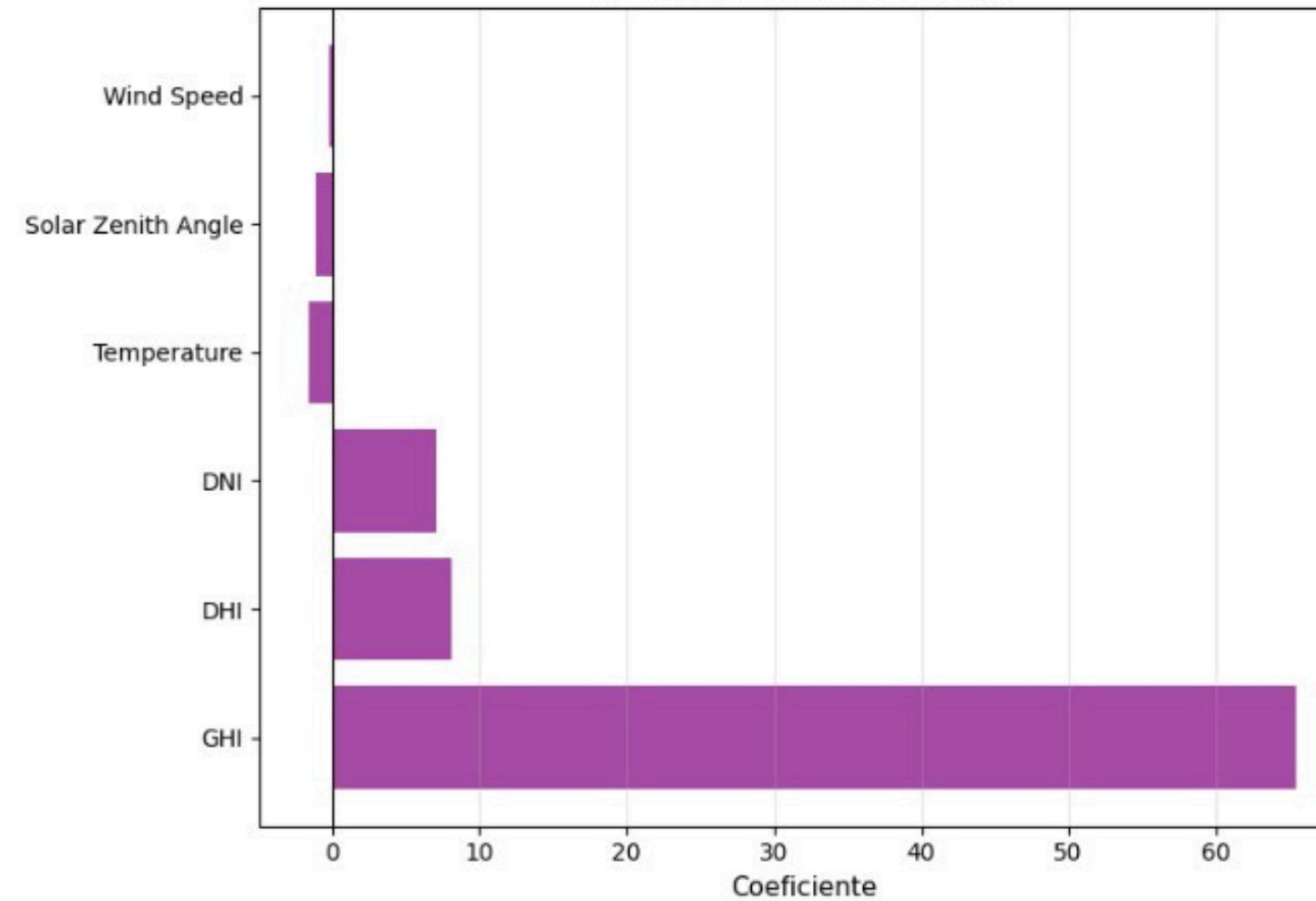
Montería - Análisis de Residuos



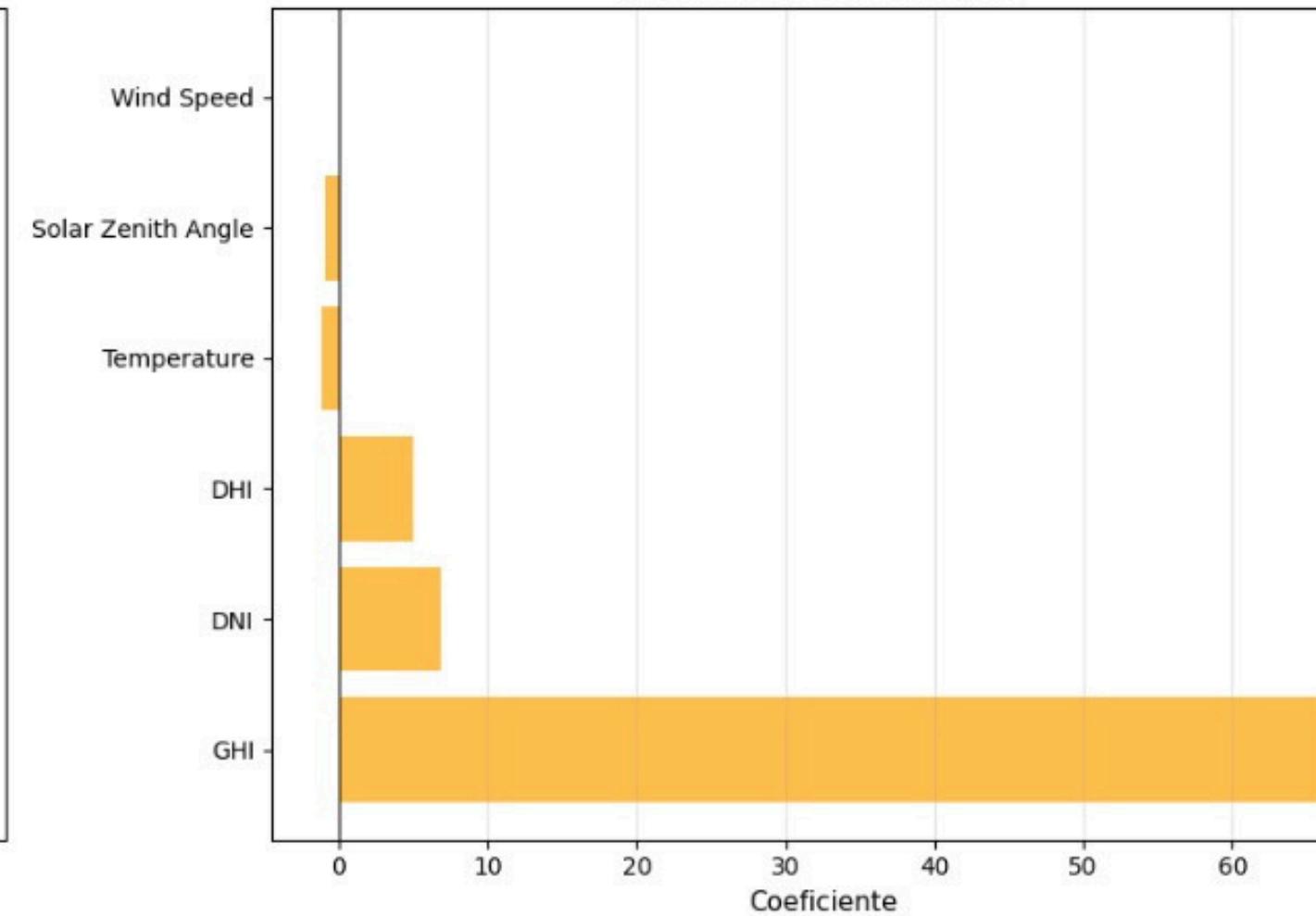
Sidney - Análisis de Residuos



**Montería - Importancia de Variables
(Coeficientes del Modelo)**



**Sidney - Importancia de Variables
(Coeficientes del Modelo)**



COEFICIENTES DEL MODELO:

Montería:

GHI	: +65.489174
DNI	: +7.069276
DHI	: +8.134893
Temperature	: -1.544263
Wind Speed	: -0.177830
Solar Zenith Angle	: -1.075865

Sidney:

GHI	: +66.213507
DNI	: +6.846583
DHI	: +5.046678
Temperature	: -1.144804
Wind Speed	: +0.073928
Solar Zenith Angle	: -0.854516

COMPARACIÓN MONTERÍA VS SIDNEY

Similitudes:

- R^2 comparable (diferencia < 1%)
- Estructura de coeficientes similar
- GHI es la variable más importante en ambos

Diferencias:

- Coeficientes ligeramente distintos
- Reflejan condiciones climáticas locales
- MAE diferente por rangos de producción distintos

CONCLUSIONES

1. Regresión lineal múltiple es altamente efectiva ($R^2 > 0.99$)
2. GHI es la variable más influyente en producción solar
3. Temperatura afecta negativamente la eficiencia
4. Modelos aplicables a diferentes ubicaciones geográficas

ANÁLISIS COMPARATIVO:
1. PRECISIÓN DEL MODELO (R^2): Montería: 0.999607 (99.96%) Sidney: 0.999133 (99.91%) → Mejor ajuste: Montería
2. ERROR ABSOLUTO MEDIO (MAE): Montería: 1.0459 Wh Sidney: 1.4721 Wh → Menor error: Montería
3. VARIABLES MÁS INFLUYENTES: Montería (por magnitud de coeficiente): 1. GHI: +65.489174 2. DHI: +8.134893 3. DNI: +7.069276 Sidney (por magnitud de coeficiente): 1. GHI: +66.213507 2. DNI: +6.846583 3. DHI: +5.046678
4. DIFERENCIAS ENTRE MODELOS: Diferencia en R^2 : 0.000474 Diferencia en MAE: 0.4262 Wh