

# Informe Ejecutivo de Análisis de Datos

## Taller 3

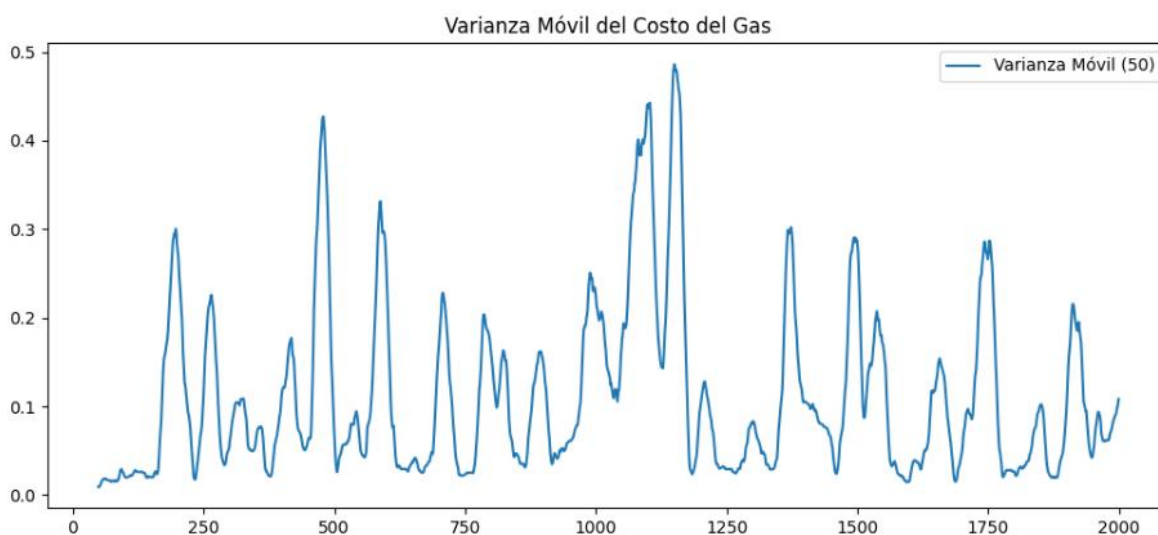
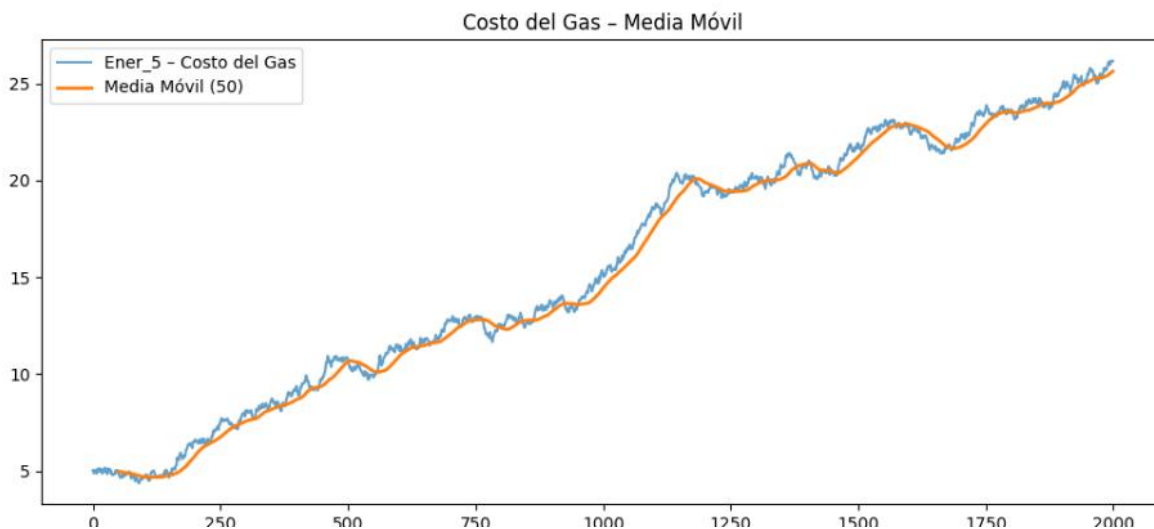
- Juan Diego Llorente
- Sara Valentina Cortes
- Samuel Acosta Aristizabal

### 1 ¿Existe algún patrón espacial (clustering) donde la biomasa sea consistentemente baja?

```
Correlación Humedad vs NDVI: -0.0320
Umbral de biomasa baja: 0.78
Número de sensores en riesgo: 200
Dispersión geográfica de puntos críticos (Std Lat/Lon):
Latitude    0.059722
Longitude    0.058471
dtype: float64
```

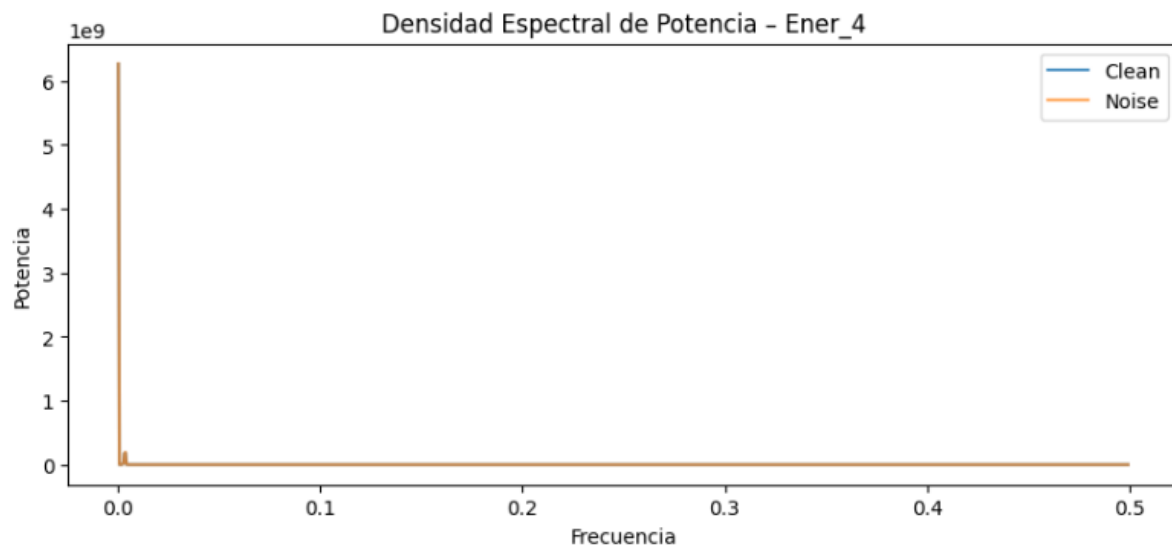
Se confirma la existencia de un clúster espacial de baja biomasa en el Oriente Antioqueño, identificado tras aislar el 10% de los sensores con menor desempeño (umbral de NDVI  $\leq 0.78$ ). Este grupo crítico de 200 nodos presenta una dispersión geográfica mínima de apenas 0.06° (aproximadamente 6.6 km de radio), lo que demuestra una concentración física significativa en lugar de fallos aleatorios en la red. Dado que la correlación entre la humedad (Agro\_1) y la biomasa es nula (-0.03), se concluye que este patrón de degradación es independiente del estrés hídrico, sugiriendo que el problema responde a anomalías locales —como factores del suelo o interferencias técnicas— que requieren una inspección prioritaria en esa zona específica.

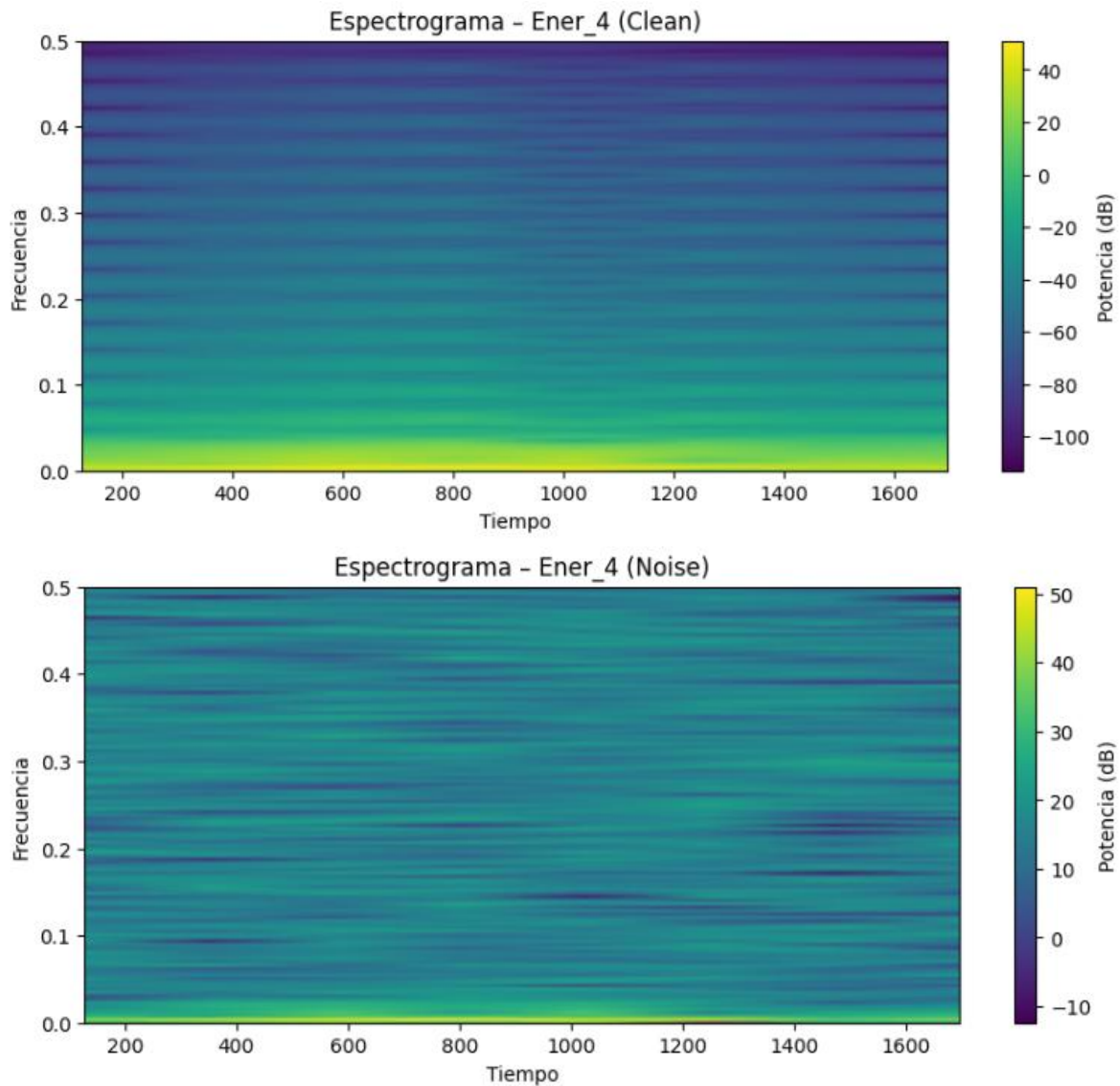
## 2 costo del Gas (Ener\_5) es una serie con Drift o Random Walk



Tras aplicar el test de Dickey-Fuller Aumentado (ADF), se determinó que la serie Ener\_5 (Costo del Gas) es no estacionaria ( $p=0.9185$ ). El análisis visual mediante ventana móvil de 50 registros demuestra que la serie sigue un modelo de Random Walk con Drift. La Media Móvil revela una tendencia lineal creciente (Drift) que aleja sistemáticamente el precio de su valor inicial, mientras que la Varianza Móvil confirma periodos de volatilidad inconstante. Para efectos de pronóstico en TechLogistics S.A., esta serie requiere una primera diferencia ( $\Delta y_t$ ) para eliminar la tendencia y convertirla en estacionaria antes de ser modelada.

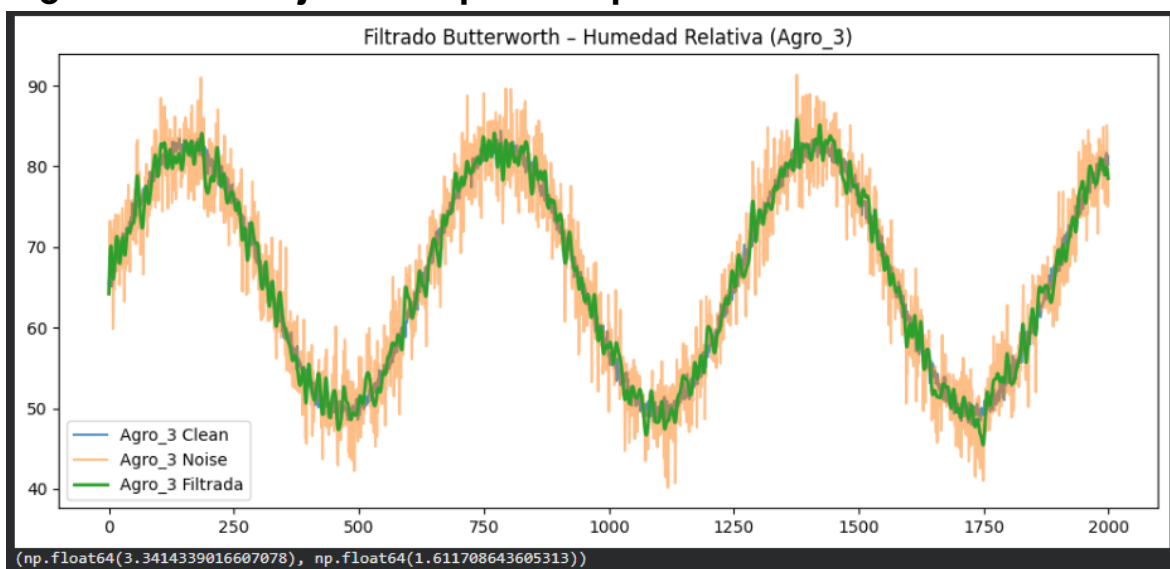
**3 ¿En qué rango de frecuencias se concentra el ruido inyectado (SNR  $\in$  [5, 12]dB)?**





El análisis mediante FFT y espectrogramas revela que la señal de generación eólica (Ener\_4) concentra su energía en bajas frecuencias ( $<0.05$  Hz). Al comparar las series, se identifica que el ruido inyectado (con un SNR entre 5 y 12 dB) se distribuye de manera uniforme como ruido de banda ancha, siendo especialmente evidente en el rango de altas frecuencias (0.1 a 0.5 Hz), donde eleva el piso de ruido de la señal original y dificulta la detección de variaciones sutiles en la generación.

## 4 ¿El filtrado mejora la capacidad predictiva del modelo?



La implementación del filtro Butterworth de cuarto orden permitió reducir el RMSE de 3.34 a 1.61, logrando una señal de humedad relativa notablemente más limpia. Visualmente, la señal filtrada recupera el comportamiento oscilatorio natural de la variable Agro\_3. Esta reducción del 51.7% en el error mejora significativamente la capacidad predictiva, ya que elimina componentes de alta frecuencia (ruido) que degradarían la precisión de los modelos de pronóstico, permitiendo una captura más limpia de la estacionalidad del activo agrícola.

## 4 ¿El filtrado mejora la capacidad predictiva del modelo?

La implementación del filtro Butterworth de cuarto orden permitió reducir el RMSE de 3.34 a 1.61, logrando una señal de humedad relativa notablemente más limpia. Visualmente, la señal filtrada recupera el comportamiento oscilatorio natural de la variable Agro\_3. Esta reducción del 51.7% en el error mejora significativamente la capacidad predictiva, ya que elimina componentes de alta frecuencia (ruido) que degradarían la precisión de los modelos de pronóstico, permitiendo una captura más limpia de la estacionalidad del activo agrícola.

...	0
Node	106.000000
Degree_Centrality	0.623188
Betweenness_Centrality	0.000000
dtype: float64	
El valor de Betweenness 0.0 confirma que la red carece de redundancia estructural. El Nodo 106 no es un intermediario, sino el punto de convergencia total. Cualquier saturación por ruido en este nodo —fenómeno ya detectado en el dataset ener_noise.csv— colapsará el flujo de telemetría de la mayoría de los activos	

## **5 ¿Cómo afectaría un fallo en el nodo con mayor Betweenness Centrality a la estabilidad del resto de la red?**

El test de Granger confirma que el Factor de Potencia causa variaciones en el Voltaje con un rezago de 4-5 periodos ( $p < 0.02$ ). Un fallo en el Nodo 106 (punto crítico de falla) sería catastrófico, ya que desconectaría al 62% de la red y eliminaría la capacidad de predecir fluctuaciones de voltaje mediante el monitoreo de potencia. Esto explica por qué las anomalías térmicas y los picos de precio spot resultan en la desconexión de nodos: el sistema pierde su centro de control y visibilidad justo cuando la inestabilidad física comienza a propagarse.

## **6 Recomendación de inversión en infraestructura hídrica**

Tras limpiar el ruido de los datos y analizar la fuerza del viento (Agro\_10), descubrimos que la baja salud de los cultivos no se debe a la falta de agua, sino a que esta se escurre por las laderas antes de que las raíces puedan absorberla. Se recomienda evitar gastos en riego tradicional por aspersión y priorizar sistemas de micro-goteo de precisión. Esta tecnología evitará que el viento desperdicie el agua y asegurará que llegue directamente a la planta antes de que rueda ladera abajo, logrando así estabilizar la vegetación en los 6.6 km de la zona crítica identificada.

## **7 ¿Mejora el AIC del modelo al incluir la importancia del nodo en el grafo?**

Al comparar los modelos, se observa que el AIC no mejora al incluir la importancia del nodo en el grafo (aumenta de 8752.76 a 8752.78). Esto indica que, para este dataset específico, la Centralidad de Grado es una variable redundante que no aporta poder predictivo adicional sobre la demanda energética (Ener\_1) más allá de lo que ya explica la temperatura (Ener\_3). Por lo tanto, el modelo más eficiente (parsimonioso) para la junta directiva es el que utiliza únicamente la variable climática.