

ANÁLISIS Y PREDICCIÓN DE LA TRANSICIÓN ENERGÉTICA EN COLOMBIA USANDO MACHINE LEARNING

INTEGRANTES

ALEXANDER TORRES VIVEROS
CAMILO ARCINIEGAS FORERO
CARLOS JAVIER CEPEDA
DAVID HERNANDO SALAMANCA
DEIBI BASTIDAS CERON

UNIVERSIDAD ICESI
FACULTAD BARBIERI DE INGENIERÍA, DISEÑO Y CIENCIAS APLICADAS
MAESTRÍA EN INTELIGENCIA ARTIFICIAL

Taller: Análisis Exploratorio y Modelos de Referencia en Energías Renovables en Colombia:

1. Análisis exploratorio de datos

El presente análisis se enfoca en el estudio de la capacidad instalada de energías renovables en Colombia, con especial énfasis en las fuentes solar, eólica y pequeñas centrales hidroeléctricas (PCH). Los datos provienen del Ministerio de Minas y Energía y fueron complementados con fuentes oficiales, entre ellas el boletín:

“Colombia supera 3 gigavatios en energías limpias dentro de su matriz energética” (MinEnergía, 2025)
<https://www.minenergia.gov.co/es/sala-de-prensa/noticias-index/colombia-supera-3-gigavatios-energias-limpias-matriz-energetica/>.

El conjunto de datos incluye registros temporales de capacidad instalada en megavatios (MW), los cuales representan el crecimiento acumulativo de las diferentes fuentes de generación limpia en el país. La serie temporal fue preprocesada mediante la conversión de formatos numéricos, eliminación de caracteres no válidos y ordenamiento cronológico, garantizando la consistencia temporal y la comparabilidad entre tecnologías.

Durante la exploración inicial se identificaron algunas inconsistencias menores —por ejemplo, valores atípicos en la variable Solar y ligeros descensos en los registros acumulativos— que podrían atribuirse a procesos de actualización o recalibración de datos en los sistemas de reporte. Dichas observaciones fueron tratadas mediante técnicas estadísticas de detección de valores extremos (IQR) y revisión de tendencias, evitando afectar la integridad del conjunto de información.

La selección de variables se fundamenta en que las tres fuentes (Solar, Eólica y PCH) representan la base de la transición energética en Colombia hacia una matriz más sostenible y diversificada. Aunque el análisis predictivo principal se aplicó sobre la variable Solar, debido a su crecimiento exponencial en los últimos años y su alto potencial en zonas como La Guajira, Atlántico y Valle del Cauca, se decidió mantener todas las variables dentro del marco analítico para permitir comparaciones cruzadas y una comprensión integral del fenómeno.

Los resultados exploratorios evidencian un incremento significativo en la capacidad instalada total, superando los 3.000 MW (3 GW) hacia 2025, lo que corresponde aproximadamente al 10 % de la matriz energética nacional. Este aumento refleja la efectividad de las políticas públicas de incentivo a las energías

limpias y la participación creciente de inversionistas privados en proyectos fotovoltaicos y eólicos a gran escala.

En términos estadísticos, las series muestran comportamientos ascendentes con variabilidad moderada; la desviación estándar de la capacidad solar es la más alta, lo cual es consistente con la rápida expansión de este tipo de proyectos en los últimos años.

En síntesis, la calidad de los datos es adecuada para análisis descriptivos y modelado predictivo, ya que cuenta con registros continuos, una estructura temporal definida y un nivel de precisión compatible con estudios energéticos a escala nacional.

2. Modelos de referencia y experimentos

Con base en la exploración anterior, se procedió a formular las hipótesis y pruebas iniciales orientadas al modelado predictivo de la capacidad de generación solar en Colombia.

Hipótesis principal (H1):

La capacidad instalada de energía solar en Colombia presenta un crecimiento sostenido de tipo exponencial, influenciado por la política nacional de transición energética y la reducción en los costos de instalación de sistemas fotovoltaicos.

Hipótesis nula (H0):

No existe una tendencia de crecimiento significativa en la capacidad instalada de energía solar durante el periodo analizado.

Para evaluar esta hipótesis, se implementaron modelos de referencia de series temporales en Python, utilizando librerías como pandas, statsmodels y scikit-learn. Se probaron enfoques lineales y suavizados para comparar comportamientos y validar la dirección de la tendencia.

En el notebook se desarrollaron los siguientes experimentos:

Visualización temporal de la capacidad solar acumulada, identificando la tendencia general ascendente y las variaciones intermensuales.

Cálculo de tasas de crecimiento y derivadas discretas (diferencias entre períodos consecutivos), para estimar la aceleración del despliegue solar.

Modelos de referencia (baseline): regresión lineal simple y ajuste polinómico de grado bajo, como punto de partida para la proyección de crecimiento futuro.

Modelo de predicción inicial: se utilizó una partición temporal de entrenamiento y prueba para evaluar la capacidad del modelo de estimar valores futuros con base en la tendencia observada.

Dentro de los modelos observados en esta etapa del proyecto se encuentran modelos de suavización exponencial, promedios móviles y ARIMA.

Primeros hallazgos encontrados después de realizar la validación con los modelos anteriormente mencionados:

1. Superioridad de Holt-Winters: El modelo Holt-Winters `mul_add_add` supera significativamente a todos los demás métodos, con un RMSE 72% menor que el mejor promedio móvil y siendo el único método avanzado factible para esta serie.
2. Limitaciones de ARIMA: Los modelos ARIMA no son adecuados para esta serie debido a la violación de supuestos fundamentales (normalidad y homocedasticidad), lo que limita su aplicabilidad práctica.
3. Evolución de la Precisión: Se observa una clara evolución en la precisión de los modelos:
 - a. Promedios móviles (método más simple): RMSE = 790.80 MW
 - b. Suavización exponencial (método intermedio): RMSE = 221.14 MW
 - c. ARIMA (método más complejo): No factible
4. Importancia de la Validación: La evaluación de supuestos estadísticos (como en ARIMA) es crucial para determinar la factibilidad de un modelo, incluso si teóricamente es más sofisticado.

El análisis experimental refuerza la hipótesis de que el país avanza consistentemente hacia una expansión acelerada de las energías limpias, en especial la solar fotovoltaica, que se proyecta como la principal fuente de generación renovable en la próxima década.

Nota: El resultado de tallado de las conclusiones se encuentra documentado en el repositorio de github.

Conclusión general

El estudio demuestra que los datos analizados reflejan con claridad el avance de Colombia en materia de generación renovable. La exploración de las variables Solar, Eólica y PCH permitió identificar la coherencia del crecimiento general y la relevancia de la energía solar como motor principal del cambio en la matriz energética.

Los modelos de referencia validados confirman que el desarrollo de la energía solar mantiene una tendencia positiva, en concordancia con los objetivos nacionales de alcanzar una mayor participación de fuentes limpias y sostenibles en el sistema eléctrico colombiano.