

EVALUACIÓN

07 AL 09 DE Noviembre

CRITERIOS DE EVALUACIÓN:

Comprensión y preparación de datos

- Calidad del análisis exploratorio y cómo se han preparado los datos para el modelado.
- Identificación de la estacionalidad y la no estacionalidad en una serie temporal.

Selección y Justificación de modelos para series de tiempo

- Utilización de gráficos como el de autocorrelación (ACF) y autocorrelación parcial (PACF).
- Selección de los parámetros p , d , y q para los modelos.

Estimación y validación del Modelo:

- Diagnóstico del modelo.
- División de datos en conjuntos de entrenamiento y prueba.
- Rendimiento del modelo utilizando métricas como MSE o MAE.

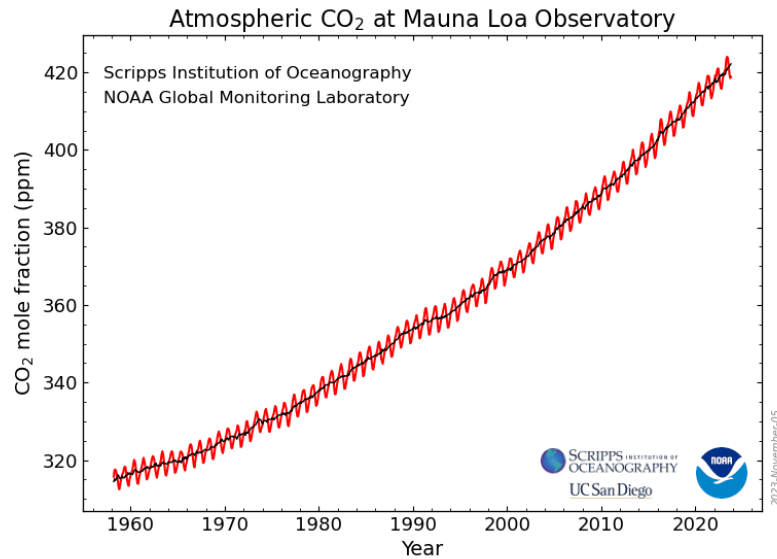
Los criterios para autoevaluación, coevaluación y heteroevaluación, se define a partir de la rúbrica anexa en la plataforma Moodle del curso

DATASET A UTILIZAR:

Primer conjunto de datos:

El dataset a utilizar lo encuentras en la siguiente página, en él se presentan los datos de CO₂ obtenidas en el emblemático Observatorio Mauna Loa en Hawái <https://gml.noaa.gov/ccgg/trends/data.html>. Este lugar es el sitio del conjunto de datos continuo más extenso de mediciones directas de CO₂ atmosférico, una serie que comenzó en 1958 bajo la dirección de Charles David Keeling del Scripps Institution of Oceanography. Este trabajo pionero forjó el camino para la comprensión moderna de las concentraciones de CO₂, y sus datos iniciales han sido continuados y ampliados por la Administración Nacional Oceánica y Atmosférica (NOAA) desde mayo de 1974.

El siguiente grafico presenta las mediciones de CO₂. incorporando la serie completa de datos que combina las mediciones originales de Scripps con las continuadas por la NOAA. Las medias mensuales que observamos se derivan de promedios diarios, los cuales, a su vez, se calculan a partir de mediciones horarias. Estas últimas se toman en cuenta únicamente cuando las condiciones atmosféricas son estables y representativas de un "fondo" no afectado por influencias locales inmediatas (para más detalles, visite gml.noaa.gov/ccgg/about/co2_measurements.html).



Segundo conjunto de datos:

Este segundo dataset es libre tomado de <https://www.datos.gov.co/> , ejemplo predecir, la temperatura, de un punto de la ciudad de Bogotá puede tomar los datos de <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Datos-Hidrometeorol-gicos-Crudos-Red-de-Estaciones/sbwg-7ju4>

```
# Instalar la biblioteca sodapy si no está instalada
!pip install sodapy

import pandas as pd
from sodapy import Socrata

client = Socrata("www.datos.gov.co", None)

results = client.get("sbwg-7ju4", limit=2000)

results_df = pd.DataFrame.from_records(results)

results_df.head()
```

DESARROLLO (para los dos dataset):

Visualización y Análisis de la Serie Temporal:

Examine la serie temporal para identificar patrones visuales, tendencias y estacionalidades. Interprete los gráficos para obtener una comprensión intuitiva de las características intrínsecas de los datos.

Prueba de Estacionariedad:

Implemente métodos estadísticamente validados, como la Prueba de Dickey-Fuller Aumentada, para evaluar si la serie temporal es estacionaria. Interprete los resultados para determinar si se requiere una transformación de los datos.

Diferenciación:

En caso de que la serie no sea estacionaria, aplique técnicas de diferenciación para estabilizar la media de la serie temporal y obtener una serie estacionaria. Justifique el número de diferenciaciones aplicadas basándose en los resultados de la prueba de estacionariedad.

Identificación de Parámetros del Modelo ARIMA:

Utilice las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los parámetros apropiados para el modelo ARIMA. Determine el número de términos autoregresivos (AR), de diferenciación (I) y de media móvil (MA) que mejor capturan la estructura de dependencia en los datos.

Validación del Modelo:

Recuerde haber dividido el conjunto de datos en dos partes: un conjunto de entrenamiento para ajustar el modelo y un conjunto de prueba para validar su desempeño. Ajuste el modelo ARIMA al conjunto de entrenamiento y evalúe su precisión predictiva en el conjunto de prueba, utilizando métricas adecuadas como el error cuadrático medio (MSE) o el error absoluto medio (MAE).

Pronóstico:

Realice predicciones futuras con el modelo ARIMA ajustado y proporcione intervalos de confianza para los pronósticos.

ENTREGABLES:

Entregables:

Proporcione un Notebook con el código fuente comentado y estructurado que haya empleado para llevar a cabo su análisis y modelado. Asegúrese de que el Notebook incluya también la siguiente información:

Visualización y Análisis de la Serie Temporal: Gráficas de la serie temporal, junto con una discusión sobre los patrones visuales, tendencias y estacionalidades identificados.

Resultados de la Prueba de Estacionariedad: Presentación de los resultados de la prueba de Dickey-Fuller Aumentada y su interpretación.

Proceso de Diferenciación: Justificación detallada del número de diferenciaciones aplicadas y la efectividad de estas para alcanzar la estacionariedad.

Selección de Parámetros del Modelo ARIMA: Tablas o gráficos de las funciones ACF y PACF, y explicación de cómo se usaron para identificar los parámetros del modelo ARIMA.

Validación del Modelo: Resumen de la metodología de división de los datos y los resultados de las métricas de desempeño utilizadas en el conjunto de prueba.

Predicción: graficas de la predicción realizada y justificación de a misma