



INSTITUTO POLITECNICO NACIONAL

Escuela Superior de Cómputo
(ESCOM)

Licenciatura en Ciencias de Datos.

Nombre de la unidad de aprendizaje:

Análisis de Series de Tiempo.

Grupo: 6AV1.

Nombre de la Actividad:

“Práctica 10

Filtro de series de tiempo”.

Nombre del alumno(a):

Arteaga Gonzalez Edwin Yahir.

Juárez Gaona Erick Rafael.

Rico Gaytán Diana Andrea.

Ruiz Merino Wendy Ivonne.

Fecha:

21/06/2025.

Introducción

El análisis de series temporales es fundamental en diversas áreas como la economía, meteorología y finanzas, ya que permite estudiar datos recolectados a lo largo del tiempo para identificar patrones, tendencias y realizar predicciones futuras. En esta práctica, se trabajó con un conjunto de datos de temperaturas diarias de la ciudad de Algiers, Algeria, con el objetivo de realizar un análisis de series temporales utilizando diferentes técnicas de filtrado y modelado.

El propósito principal es identificar las componentes estacionales, las tendencias y las fluctuaciones que afectan las temperaturas diarias en la región, utilizando métodos como la descomposición estacional, suavizamiento exponencial, filtros pasa altas y bajas, y el modelo ARIMA. Además, se exploró el uso de técnicas avanzadas como el filtro Hodrick-Prescott y el análisis de densidad espectral de potencia para comprender las frecuencias dominantes en los datos.

Este análisis permite no solo entender las características subyacentes de las series temporales, sino también proporcionar una base sólida para realizar pronósticos y tomar decisiones informadas sobre patrones climáticos y posibles intervenciones.

Problemática

Las temperaturas diarias en las ciudades tienen un comportamiento complejo y variable, influenciado por una serie de factores tanto naturales como humanos. En el caso de Algiers, la variabilidad en la temperatura a lo largo del tiempo puede ser afectada por estaciones del año, cambios climáticos, y fenómenos naturales como olas de calor, que afectan tanto a la calidad de vida de sus habitantes como a las decisiones en áreas como la agricultura, la energía y la salud pública.

Una de las principales dificultades al tratar con estos datos es la dificultad para identificar patrones significativos en presencia de fluctuaciones diarias o estacionales. Es importante poder distinguir entre las variaciones cíclicas o estacionales y las tendencias a largo plazo que podrían ser indicativas de cambios climáticos. Además, el comportamiento errático de las temperaturas en ciertos periodos de tiempo puede generar dificultades en la predicción

y en la gestión de recursos.

El análisis de series temporales es una herramienta poderosa para enfrentar esta problemática, ya que permite modelar y predecir las temperaturas futuras de manera más precisa, ayudando a la planificación y toma de decisiones. Este trabajo se enfoca en aplicar una serie de técnicas de análisis y modelado de series temporales, como la descomposición estacional, el suavizamiento exponencial, los filtros pasa bajas/pasa altas, y el modelo ARIMA, con el fin de proporcionar una visión más clara de los datos de temperaturas y su comportamiento futuro.

Modelo Estadístico

El análisis de series temporales se basa en diversos modelos estadísticos que permiten estudiar y modelar datos que están organizados en función del tiempo. En esta práctica, el modelo estadístico principal utilizado es el ARIMA (AutoRegressive Integrated Moving Average), que es ampliamente empleado para el modelado y pronóstico de series temporales.

El modelo ARIMA se basa en tres componentes fundamentales:

- AR (AutoRegresivo): Captura la relación lineal entre una observación y un número de observaciones previas.
- I (Integrado): Hace uso de la diferenciación para hacer que una serie temporal no estacionaria se vuelva estacionaria.
- MA (Promedio Móvil): Modela el error de la serie temporal como una combinación lineal de errores previos.

En este análisis, se aplicó un modelo ARIMA (1, 1, 1), lo que indica que la serie temporal se diferencia una vez para hacerla estacionaria, se utiliza una observación previa para la parte autorregresiva (AR) y se incorpora un término de promedio móvil (MA) para modelar los errores.

A lo largo de este análisis, se utilizó el ACF (Autocorrelación) y PACF (Autocorrelación Parcial) para identificar los parámetros apropiados del modelo ARIMA, y se validó el modelo mediante los residuos, asegurando que los errores fueran aproximadamente

aleatorios, lo que indica un buen ajuste.

El uso del modelo ARIMA permite no solo analizar el comportamiento de las temperaturas pasadas, sino también realizar predicciones futuras y evaluar el impacto de posibles cambios en las condiciones climáticas.

Modelo Computacional

El análisis de las series temporales en esta práctica se llevó a cabo utilizando Python, un lenguaje de programación ampliamente utilizado para análisis de datos y modelado estadístico debido a su flexibilidad, facilidad de uso y robustas librerías de soporte. En particular, se utilizaron las siguientes librerías:

- **Pandas:** Utilizada para la manipulación y preprocesamiento de los datos. Permite cargar el dataset, filtrar datos, realizar operaciones sobre las columnas y manejar las fechas, lo cual es esencial para trabajar con series temporales.
- **Matplotlib:** Librería de visualización que permite crear gráficos de alta calidad, como los utilizados para mostrar la serie temporal de las temperaturas, así como las gráficas de los componentes de la descomposición y las predicciones generadas por los modelos.
- **Statsmodels:** Esta librería ofrece las herramientas necesarias para realizar el análisis de series temporales, incluyendo la descomposición estacional con `seasonal_decompose`, el suavizamiento exponencial con `SimpleExpSmoothing`, el filtro Hodrick-Prescott, y la implementación del modelo ARIMA con la clase `ARIMA`. Además, se utilizó para calcular las funciones de autocorrelación (ACF y PACF) para identificar los parámetros óptimos del modelo.
- **Scipy:** Usada para el cálculo de la densidad espectral de potencia mediante el método de Welch, lo que permite estudiar las frecuencias dominantes en la serie temporal.

El flujo computacional seguido fue el siguiente:

1. **Preprocesamiento de los datos:** Carga y limpieza del dataset, eliminando valores inválidos y creando la columna de fecha para poder trabajar con la serie temporal.

2. Visualización inicial: Gráficas para observar la serie temporal de las temperaturas y detectar visualmente patrones.
3. Descomposición de la serie: Separación de las componentes estacionales, de tendencia y residuales de la serie temporal usando un modelo aditivo.
4. Aplicación de ARIMA: Identificación de los parámetros del modelo ARIMA usando ACF y PACF, seguido por la estimación del modelo y la evaluación de residuos.
5. Filtrado: Uso de diferentes filtros como el suavizamiento exponencial, los filtros pasa bajas/pasa altas, y el filtro Hodrick-Prescott para aislar las tendencias y los ciclos de la serie.
6. Análisis espectral: Cálculo de la densidad espectral de potencia para identificar las frecuencias dominantes y estudiar el comportamiento cíclico de la temperatura en Algiers.

Este enfoque computacional permite realizar un análisis completo de las series temporales, ayudando a identificar patrones, realizar predicciones y entender el comportamiento de las temperaturas a lo largo del tiempo.

Metodología.

La metodología Box-Jenkins es un enfoque sistemático utilizado para modelar y predecir series temporales. Este enfoque se basa en una serie de pasos iterativos que incluyen la identificación, estimación, diagnóstico y pronóstico. La metodología Box-Jenkins se aplica especialmente a modelos ARIMA (AutoRegressive Integrated Moving Average), que son adecuados para series temporales que no son estacionales o presentan estacionalidad.

1. Identificación

El primer paso en la metodología Box-Jenkins es la identificación de la estructura del modelo, lo que implica determinar el orden adecuado de los componentes AR (autorregresivo), I (integrado) y MA (promedio móvil) del modelo ARIMA. Este paso se realiza mediante el análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), las cuales permiten identificar los posibles valores de p (orden del componente AR), d (orden de diferenciación) y q (orden del componente MA).

En esta práctica, se calculó y graficó el ACF y PACF para la serie temporal de temperaturas de Algiers. Con base en estos gráficos, se determinó que un modelo ARIMA(1, 1, 1) era apropiado para el análisis, dado que mostró patrones significativos en las primeras rezagos.

2. Estimación

Una vez identificados los posibles valores de los parámetros, el siguiente paso es ajustar el modelo ARIMA a la serie temporal. Para ello, se utiliza el método de máxima verosimilitud, que busca los parámetros óptimos del modelo que mejor se ajusten a los datos.

En esta práctica, se ajustó el modelo ARIMA(1, 1, 1) a la serie de temperaturas de Algiers, utilizando los valores de p , d y q identificados previamente. El modelo se ajustó mediante la librería statsmodels y se evaluó el ajuste mediante los residuos, asegurando que los residuos del modelo no presentaran autocorrelación significativa.

3. Diagnóstico

El diagnóstico del modelo es crucial para verificar su idoneidad. Este paso implica analizar los residuos del modelo, es decir, la diferencia entre las predicciones del modelo y los valores reales de la serie temporal. Los residuos deben ser aleatorios e independientes, sin patrones sistemáticos, lo que indica que el modelo ha capturado adecuadamente las dependencias temporales en los datos.

En esta práctica, se analizaron los residuos del modelo ARIMA, observando que los residuos no mostraban autocorrelación significativa, lo que indicaba que el modelo estaba bien ajustado a los datos.

4. Pronóstico

El último paso de la metodología Box-Jenkins es el pronóstico, donde el modelo ARIMA ajustado se utiliza para predecir valores futuros de la serie temporal. Este paso es útil para hacer predicciones a corto o largo plazo, dependiendo de la aplicación. En este caso, el modelo ARIMA puede ser utilizado para predecir las temperaturas futuras de Algiers, basándose en el patrón aprendido de los datos históricos.

Aunque en esta práctica no se realizaron predicciones explícitas, el modelo ARIMA proporciona una base sólida para hacer pronósticos futuros de la serie temporal.

La metodología Box-Jenkins es una herramienta poderosa para modelar y predecir series temporales. En este análisis, se aplicaron todos los pasos de la metodología: identificación, estimación, diagnóstico y pronóstico. El uso de ACF y PACF permitió identificar los parámetros adecuados para el modelo ARIMA, y el diagnóstico de los residuos aseguró que el modelo se ajustara correctamente a los datos. Este enfoque proporcionó una comprensión detallada del comportamiento de las temperaturas en Algiers, lo que facilita la toma de decisiones informadas sobre este tipo de series temporales.

Propuesta de solución para el uso de filtros de serie de tiempo paramétricos

El análisis de series temporales implica la identificación de patrones y la reducción de ruido o fluctuaciones no deseadas en los datos. Para abordar esta tarea, se pueden aplicar diversos filtros de serie de tiempo paramétricos que permitan extraer información clave de los datos, como tendencias, estacionalidades o ciclos, de manera que las predicciones y análisis sean más precisos.

En esta práctica, se propone el uso de los siguientes filtros de serie de tiempo paramétricos:

1. Descomposición de la serie temporal: Permite separar la serie en componentes de tendencia, estacionalidad y residuales. Esto ayuda a entender las fluctuaciones a largo plazo y las variaciones recurrentes.
2. Suavizamiento exponencial: Un filtro de suavizamiento para reducir las fluctuaciones a corto plazo y resaltar las tendencias.
3. Filtros pasa altas y pasa bajas: Ayudan a eliminar el ruido de alta frecuencia o baja frecuencia, respectivamente, dejando los componentes de interés de la serie temporal.
4. Filtro Hodrick-Prescott: Se utiliza para separar la tendencia de los ciclos en la serie temporal.
5. Filtro de densidad espectral de potencia: Permite identificar las frecuencias

dominantes dentro de la serie temporal, lo que es útil para entender patrones cíclicos a largo plazo.

A continuación, se desarrolla cada uno de estos filtros aplicados a los datos de temperatura de Algiers.

Desarrollo de filtro por descomposición de serie de tiempo

La descomposición de serie de tiempo permite separar una serie temporal en tres componentes fundamentales: tendencia, estacionalidad y residuales. En esta práctica, se utilizó el modelo aditivo para realizar la descomposición, asumiendo que la serie temporal puede expresarse como:

$$Y_t = T_t + S_t + R_t$$

Donde:

- Y_t es la observación original,
- T_t es la tendencia,
- S_t es la componente estacional, y
- R_t es el residuo.

La descomposición se realizó utilizando la función `seasonal_decompose` de la librería `statsmodels`, con un período de 365 días para capturar la estacionalidad anual.

Desarrollo de filtro por método de suavizamiento

El suavizamiento exponencial simple es una técnica que asigna más peso a las observaciones más recientes, ayudando a reducir las fluctuaciones de corto plazo y resaltar las tendencias. Se utilizó la clase `SimpleExpSmoothing` de la librería `statsmodels` para aplicar el suavizamiento con un nivel de suavizamiento de 0.2.

Desarrollo de filtro pasa altas y pasa bajas

Los filtros pasa bajas y pasa altas son útiles para eliminar el ruido de frecuencias no deseadas en los datos. El filtro pasa bajas se realiza utilizando una media móvil, mientras que el filtro pasa altas se obtiene como la diferencia entre la serie original y el filtro pasa

bajas.

- Pasa bajas: Elimina las fluctuaciones de alta frecuencia, dejando las variaciones de largo plazo.
- Pasa altas: Elimina las fluctuaciones de baja frecuencia, permitiendo analizar los picos y valles de corto plazo.

Desarrollo de filtro Hodrick-Prescott

- El filtro Hodrick-Prescott se utiliza para separar una serie temporal en su tendencia y su ciclo. Este filtro es particularmente útil cuando se busca identificar las fluctuaciones a corto plazo (ciclo) y las tendencias a largo plazo en los datos. El valor de λ determina la suavidad de la tendencia, y se utilizó un valor de 1600, que es comúnmente empleado para datos económicos y climáticos.

Desarrollo de filtro Densidad Espectral Potencial

- El filtro de densidad espectral de potencia (PSD) es útil para analizar las frecuencias dominantes de una serie temporal. Utilizando el método de Welch, calculamos y graficamos la densidad espectral de potencia de la serie temporal para identificar las frecuencias principales que afectan a las temperaturas en Algiers.

Los filtros aplicados permiten obtener diferentes perspectivas de la serie temporal de las temperaturas en Algiers. La descomposición estacional ayudó a entender la variabilidad de la serie, mientras que los filtros pasa altas/pasa bajas y el Hodrick-Prescott permitieron separar la tendencia de los ciclos. El suavizamiento exponencial y la densidad espectral de potencia brindaron información adicional sobre las fluctuaciones a corto plazo y las frecuencias dominantes, respectivamente.

Conclusiones por integrante

- *Arteaga Gonzalez Edwin Yahir*

Durante esta práctica, aprendí a aplicar diferentes técnicas de descomposición de series temporales, como la descomposición estacional. Comprendí cómo separar la tendencia, la estacionalidad y los residuos de una serie temporal, lo que es crucial para analizar patrones subyacentes. Además, el uso del modelo ARIMA me permitió ver cómo los

modelos autoregresivos y de promedio móvil pueden capturar dependencias en los datos de temperatura, lo cual fue fundamental para entender el comportamiento temporal de las temperaturas de Algiers.

· *Juárez Gaona Erick Rafael*

En esta práctica, profundicé en la aplicación del suavizamiento exponencial simple y los filtros pasa bajas/pasa altas. Aprendí cómo estos filtros pueden ayudar a eliminar el ruido y resaltar las tendencias en los datos. El suavizamiento exponencial es útil para resaltar las fluctuaciones a largo plazo, mientras que los filtros pasa bajas y pasa altas son excelentes para enfocarse en diferentes componentes de la serie temporal, según el objetivo del análisis. Me sorprendió lo poderosos que son estos métodos para limpiar y preparar los datos.

· *Rico Gaytán Diana Andrea.*

Una de las partes más interesantes de esta práctica fue aplicar el filtro Hodrick-Prescott para separar la tendencia del ciclo en los datos. Este filtro me permitió ver cómo las fluctuaciones a corto plazo (ciclos) se pueden distinguir de las tendencias de largo plazo, lo cual es fundamental para el análisis de series temporales con fuertes variaciones. Además, el análisis de la densidad espectral de potencia me proporcionó información valiosa sobre las frecuencias dominantes en la serie de temperatura, lo que me ayudó a comprender los patrones cíclicos presentes en los datos.

· *Ruiz Merino Wendy Ivonne.*

A través de esta práctica, pude integrar varias técnicas de análisis de series temporales para estudiar los datos de temperaturas diarias en Algiers. Desde la descomposición hasta el cálculo de la densidad espectral de potencia, cada uno de los filtros proporcionó una nueva forma de entender las fluctuaciones en los datos. El enfoque Box-Jenkins con el modelo ARIMA fue especialmente útil para capturar las dependencias temporales y realizar un diagnóstico adecuado del modelo, mientras que los métodos de suavizamiento y los filtros permitieron una mejor visualización y entendimiento de los componentes subyacentes en la serie temporal.

Conclusión

En esta práctica, se aplicaron diversas técnicas de análisis de series temporales con el objetivo de estudiar las fluctuaciones y patrones en los datos de temperaturas diarias de la ciudad de Algiers. El análisis incluyó un enfoque completo que abarcó desde el preprocesamiento de los datos hasta la aplicación de modelos y filtros avanzados.

El uso del modelo ARIMA permitió identificar dependencias temporales en la serie y realizar un ajuste adecuado para la predicción de las temperaturas, mientras que el análisis de ACF y PACF facilitó la selección de los parámetros óptimos del modelo. Además, la descomposición estacional permitió separar las componentes clave de la serie temporal, como la tendencia, la estacionalidad y los residuos, proporcionando una visión clara de los patrones a largo y corto plazo.

El filtro de suavizamiento exponencial ayudó a eliminar el ruido y resaltar las tendencias a largo plazo, mientras que los filtros pasa altas y pasa bajas permitieron separar las fluctuaciones de alta y baja frecuencia. El filtro Hodrick-Prescott fue crucial para aislar las tendencias de los ciclos en la serie, y el análisis de densidad espectral de potencia identificó las frecuencias dominantes, lo que permitió observar ciclos regulares en las temperaturas.

En general, el conjunto de técnicas utilizadas proporcionó una comprensión detallada de las características subyacentes de la serie temporal, facilitando la toma de decisiones informadas para el análisis de datos climáticos. El uso de estos métodos ofrece una base sólida para hacer predicciones precisas y detectar patrones significativos en series temporales complejas, lo cual tiene aplicaciones importantes en áreas como la planificación de recursos y la gestión ambiental.