

## **Predicción del Índice de Precios al Consumo Armonizado (HICP) en la Unión Europea**

### **Introducción**

El Índice de Precios al Consumo Armonizado (HICP por sus siglas en inglés) es una medida estandarizada utilizada por los países de la Unión Europea para calcular la inflación.

A diferencia de los índices de precios nacionales, como el IPC en el caso de España, el HICP permite comparaciones consistentes entre países, ya que se construyen siguiendo una metodología común definida por la Eurostat.

Este índice juega un papel fundamental en la política económica de la Unión Europea. Es uno de los indicadores clave utilizado por el Banco Central Europeo para tomar decisiones sobre tipos de interés y evaluar la estabilidad de precios. Asimismo, es una herramienta crítica para gobiernos, analistas y empresas que deben anticiparse a la evolución del poder adquisitivo y a las condiciones económicas de los distintos estados miembros.

Dada su relevancia, poder predecir el comportamiento futuro del HICP con precisión es de gran utilidad para múltiples agentes económicos, como los bancos centrales que ajustan la política monetaria.

Por ejemplo, una empresa de distribución minorista con operaciones en varios países europeos necesita anticipar la evolución de la inflación para ajustar sus precios, negociar contratos con proveedores y actualizar previsiones de costes logísticos. Del mismo modo, compañías que ofrecen productos financieros indexados al coste de vida, como seguros o pensiones, dependen de estimaciones fiables del HICP para proteger sus márgenes y ofrecer garantías competitivas.

En todos estos casos, disponer de modelos predictivos precisos permite a las organizaciones actuar de forma proactiva, en lugar de reaccionar ante cambios inesperados del entorno económico.

En este contexto, el proyecto tiene como objetivo diseñar un sistema de modelado y predicción del HICP para varios países de la Unión Europea, utilizando diferentes enfoques de series temporales y aprendizaje automático. Dichos países han sido seleccionados teniendo en cuenta la similitud entre ellos en cuanto a estructura y PIB.

### **Objetivo**

El propósito de este trabajo es construir y comparar distintos modelos predictivos para estimar la evolución mensual del HICP en siete países de la Unión Europea cuyas estructuras y PIB son similares: Bélgica, Italia, Francia, España, Polonia, Países Bajos y Alemania.

Se han implementado y evaluado tres enfoques distintos:

- Modelo ARIMA: adecuados para serie temporales univariantes, con o sin estacionalidad.
- Modelo SARIMAX: Permiten modelar componentes estacionales más complejos y relaciones con posibles variables exógenas.
- Modelo XGBoost: una técnica de aprendizaje automático basada en árboles, utilizada aquí como modelo autoregresivo con variables calendario y retardos (lags).

La comparación se ha realizado tanto en términos de capacidad predictiva en el corto plazo como de ajuste sobre datos reales de 2025, lo que proporciona una validación externa robusta.

### **Estructura del dataset original:**

Los datos extraídos desde la API de la Unión Europea (<https://data.europa.eu>) siguen la estructura estándar utilizada para reportar el Índice Armonizado de Precios al Consumo (HICP).

Las columnas principales incluyen:

- freq: Frecuencia temporal (mensual).
- unit: Unidad de medida, aquí "I05", es decir, índice base 2015 = 100.
- coicop: Clasificación de consumo COICOP, permite segmentar por tipo de gasto (alimentos, transporte, etc.).
- geo: País o región.
- date: Fecha (mes/año).
- hicp: Valor del índice.

Se emplea la codificación COICOP para identificar grupos de consumo. Ejemplos relevantes:

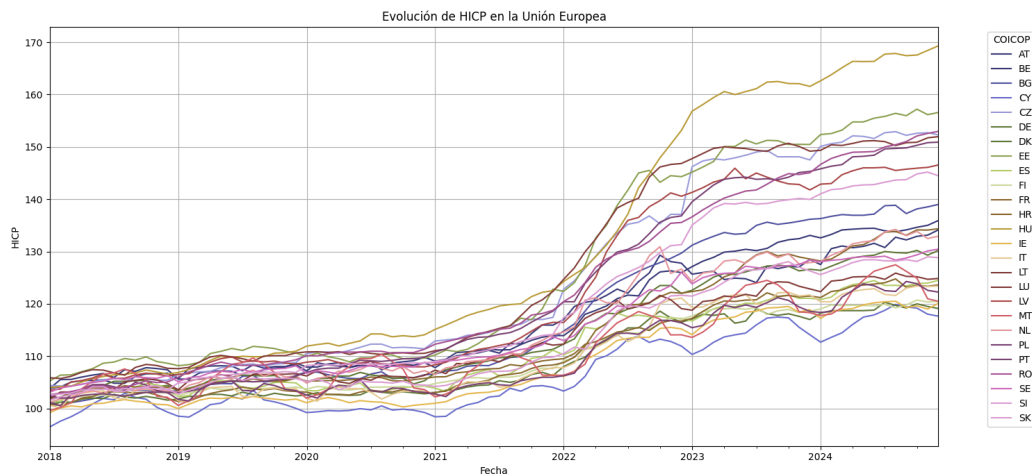
- CP00: Consumo total.
- CP01111: Pan.
- CP07: Transporte.
- TOT\_X\_FUEL: Total sin combustibles.
- SERV\_TRA: Servicios de transporte.

En este proyecto se trabajó únicamente con CP00, pero esta estructura permite futuras extensiones a categorías más específicas o combinadas.

## Datos y procesamiento

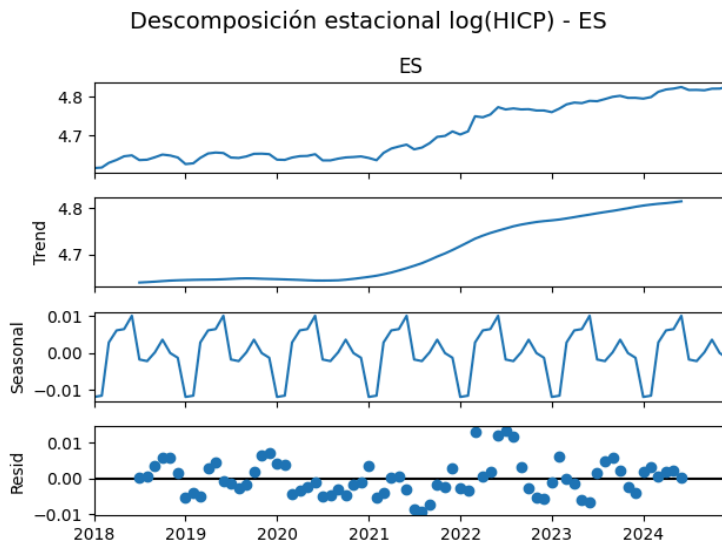
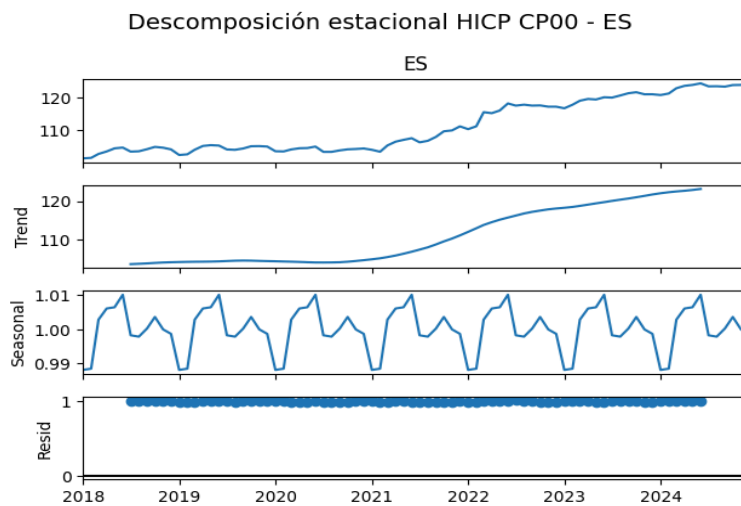
Tal y como se menciona en el apartado anterior, los datos utilizados son reales y han sido extraídos de la API de la Unión Europea. El dataset contiene la evolución del HICP por país y categoría de consumo (COICOP) desde 1996, aunque para este proyecto se ha trabajado con datos desde enero de 2018.

- Se filtró únicamente la categoría CP00 (Consumo total) y los países seleccionados.
- Las series se transformaron a formato univariante por país utilizando pivoteo.
- Se generaron variables auxiliares como mes, año y lags para alimentar los modelos no lineales (XGBoost).
- Además, se aplicaron transformaciones logarítmicas y descomposición estacional (aditiva y multiplicativa) para mejorar el rendimiento de los modelos ARIMA.



## Análisis exploratorio y descomposición

En primer lugar se realizó una exploración visual de la evolución del HICP en los países seleccionados. Las series muestran una clara tendencia creciente y patrones estacionales marcados. Para un mayor estudio de dicha estacionalidad, se aplicó una descomposición estacional aditiva y multiplicativa para separar los componentes de tendencia, estacionalidad y ruido. Esto ha permitido analizar con mayor precisión la dinámica interna de cada serie y facilitó el desarrollo de modelos entrenados exclusivamente en la componente de tendencia.



*Se aplicó una transformación logarítmica para estabilizar la varianza y permitir una separación más clara de la estacionalidad.*

### Modelos predictivos:

Arima y Sarimax:

Se entrenaron estos dos modelos para cada país, tanto sobre la serie completa como sobre la componente de tendencia. Los modelos fueron seleccionados automáticamente con AutoArima, optimizando los parámetros  $p$ ,  $d$ ,  $q$  y la estacionalidad ( $m=12$ ).

Se validó la calidad del ajuste analizando los residuos con ACF, PACF y mediante el test Ljung-Box para comprobar que fueran ruido blanco, es decir que no presenta autocorrelación (no hay dependencia entre los valores pasados y futuro, con media cero y varianza constante, y es completamente impredecible). Como resultado, si el modelo ha captado bien toda la estructura de la serie (tendencias, estacionalidad, etc.), entonces los residuos son ruido blanco, si los residuos no son blancos, por el contrario indicaría que queda información estructurada sin modelar y el modelo podría mejorarse

XGBoost:

Se utilizó como un modelo supervisado, alimentado con variables temporales (mes y año) y valores pasados del HICP (lags). Se entrenó por país, evaluando su capacidad para aprender patrones no lineales de las series temporales.

### Evaluación y comparación

Los modelos fueron evaluados usando:

- Root Mean Squared Error
- Mean Absolute Error

Primero se compararon los modelos en test internos (hasta diciembre de 2024), y luego se realizó una validación externa con datos reales del 2025:

[https://ec.europa.eu/eurostat/databrowser/view/prc\\_hicp\\_midx/default/table?lang=en&category=prc.prc\\_hicp](https://ec.europa.eu/eurostat/databrowser/view/prc_hicp_midx/default/table?lang=en&category=prc.prc_hicp)

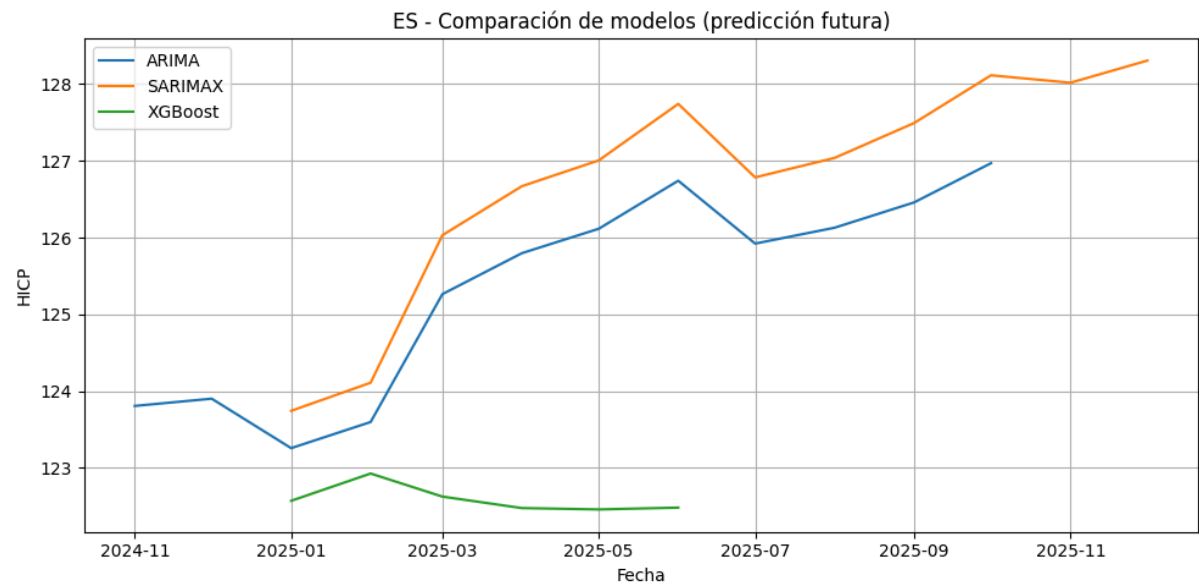
País	Modelo	RMSE	MAE
BE	ARIMA	1.821212352188205	1.5895796300247014
BE	SARIMAX	0.8213656437112579	0.6447471074663819
BE	XGBoost	2.4349015505461526	2.3120554606119774
DE	ARIMA	1.0089556251188463	1.0049486319861007
DE	SARIMAX	0.8836936310225855	0.8270094587523621
DE	XGBoost	1.0153752510620353	1.0029266357421893
ES	ARIMA	0.9514598539380278	0.8846151404523775
ES	SARIMAX	0.5589373178304275	0.5431427058033526
ES	XGBoost	2.2915491297458797	2.2165589396158842
FR	ARIMA	0.5427736556061091	0.4698960116703432
FR	SARIMAX	0.9114873807972317	0.7539229669254155
FR	XGBoost	0.23358143223957184	0.22401814778646192
IT	ARIMA	0.525668786975979	0.4860157819936101
IT	SARIMAX	0.5552592076526452	0.5394957224991259
IT	XGBoost	1.6675733412512694	1.5611638387044309

NL	ARIMA	1.6736030093958232	1.5374097048413187
NL	SARIMAX	0.42819224825912855	0.2964441144266914
NL	XGBoost	0.814573098701955	0.7115004475911443
PL	ARIMA	0.4859796943822241	0.4830606656494467
PL	SARIMAX	0.4703171200785175	0.312288091663343
PL	XGBoost	2.0292555687573506	1.9875732421874943

**Conclusiones:**

El desempeño varió notablemente por país, lo que refuerza la decisión de utilizar un modelo específico por país.

- SARIMAX fue el modelo más robusto en términos generales, especialmente en países con fuerte estacionalidad.
- XGBoost resultó competitivo e incluso superior en algunos contextos, mostrando su capacidad para capturar patrones complejos.
- ARIMA sigue siendo una solución simple y eficaz cuando los patrones son más regulares.



**Líneas a seguir desarrollando**

- Incorporar variables exógenas (indicadores macroeconómicos)
- Extender el análisis a más países y categorías COICOP

- Desplegar los modelos como dashboard interactivos como Streamlit.