BANCO CENTRAL DE COSTA RICA

JORNADAS DE INVESTIGACIÓN ECONÓMICA 2026

Métodos de aprendizaje automático para la predicción de la inflación en múltiples horizontes: **un análisis para Costa Rica**

Autor: Manuel Esteban Sánchez Gómez

17 de noviembre, 2026

BANCO CENTRAL DE COSTA RICA

JORNADAS DE INVESTIGACIÓN ECONÓMICA 2026

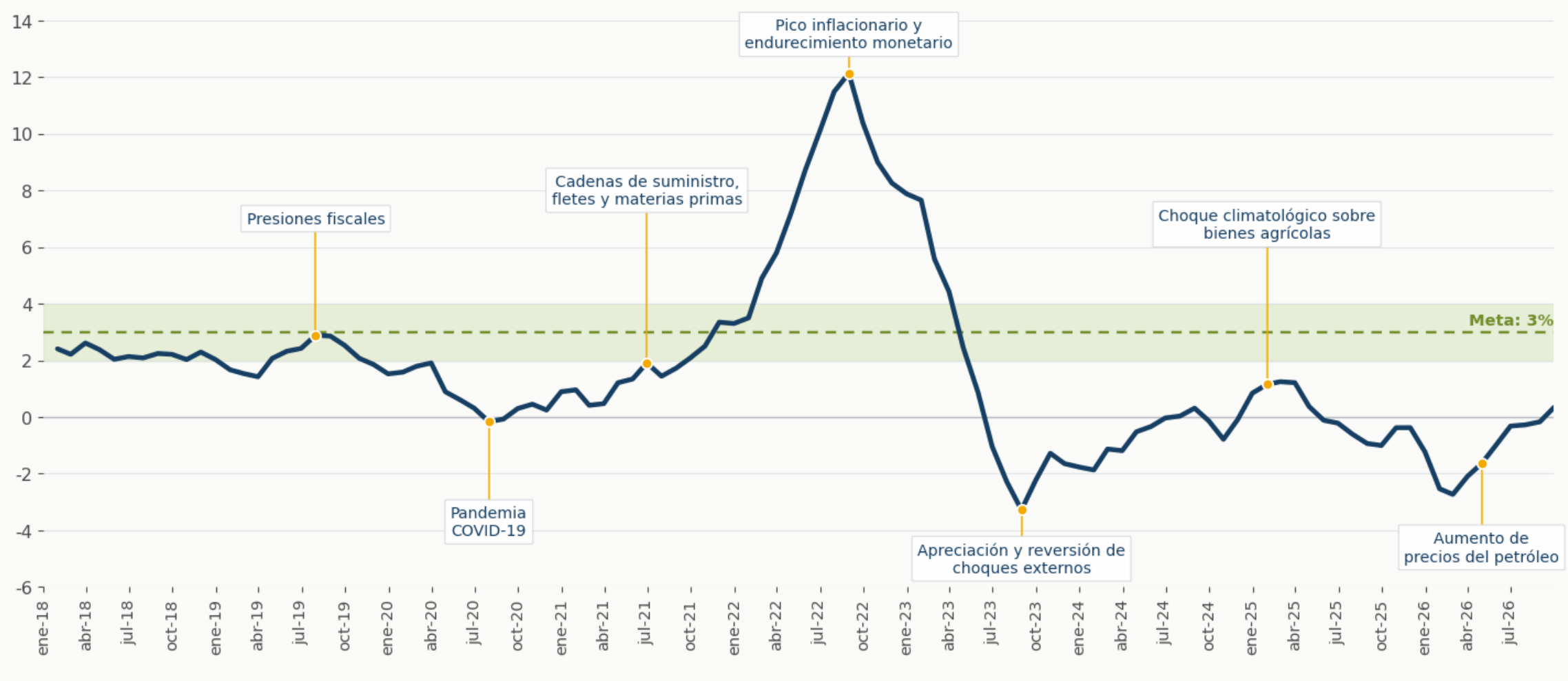
Las ideas aquí expresadas son del autor y no necesariamente representan las del Banco Central de Costa Rica.

Tenemos más información, pero pronosticar es aún una labor compleja…

La última década combina eventos de naturaleza muy distinta y relaciones que cambian en el tiempo.

Inflación general

Variación interanual en porcentajes



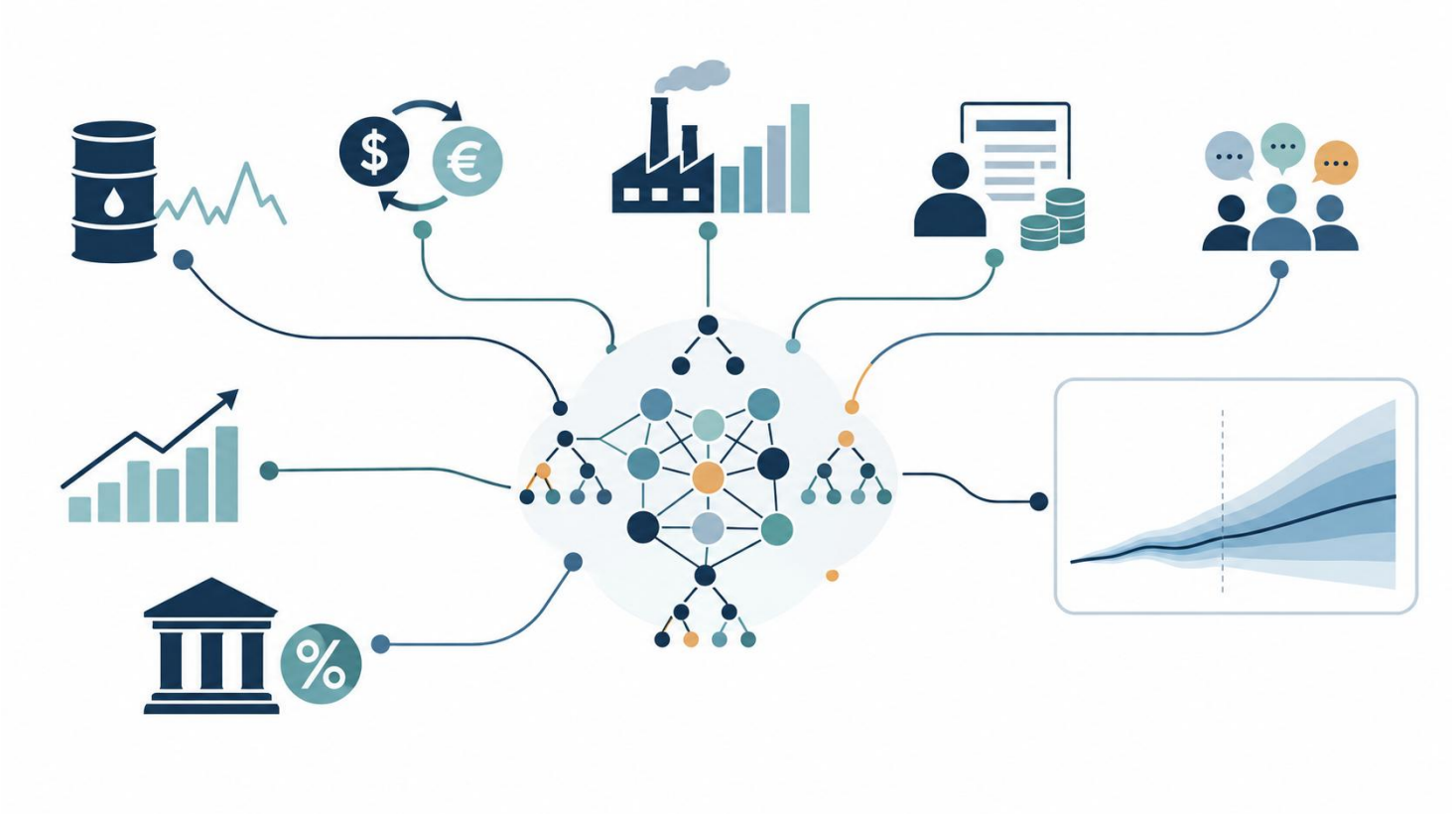
Fuente: Elaboración propia con base en datos del INEC.

¿Qué pueden aportar los modelos de aprendizaje automático?

***Machine Learning*: métodos estadísticos y computacionales que utilizan los datos para identificar patrones y relaciones complejas útiles para predecir.**

INTERÉS PARA LA DIVISIÓN ECONÓMICA

Los pronósticos de corto plazo alimentan modelos más estructurales y apoyan el análisis de la inflación y las recomendaciones de política.



Alta dimensionalidad

Podemos procesar muchas señales simultáneamente sin imponer una ecuación única.

Interacciones

una variable puede tomar más relevancia cuando otra también se mueve.

No linealidades

El efecto puede cambiar según la magnitud o el estado de la economía.

Complemento a otros modelos

Incrementa la cantidad de herramientas disponibles para el análisis de la inflación.

LA INVESTIGACIÓN

El experimento original

Una competencia controlada: misma información, mismo protocolo, distintos modelos.

43 series

Precios, actividad, expectativas, mercado laboral, finanzas y condiciones monetarias.

172 predictores

Rezagos 1, 3, 6 y 12 meses.

Ajuste de acuerdo con la disponibilidad de datos.

Varios modelos

Lineales

No lineales

Híbridos

Variable dependiente: **Inflación interanual**

Varios horizontes

Pronósticos para:

H+1

H+3

H+6

H+12

Validación

Entrenamiento: 2010 - 2023.

Evaluación: 2024–2025.

Calibración de hiperparámetros · Ventanas móviles · *Benchmarks*· RMSE/MAE · Diebold–Mariano

Objetivo: comparar capacidad predictiva para el caso de Costa Rica.

LA INVESTIGACIÓN

Muchos algoritmos tratan de resolver un mismo problema

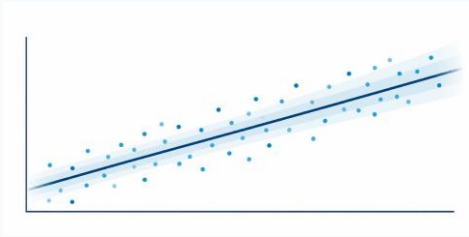
El ejercicio compara enfoques distintos para una misma pregunta: ¿qué modelo extrae mejor la señal útil para anticipar la inflación?

Lineales regularizados

Ridge · Lasso · Elastic Net

Buscan una relación lineal, pero penalizan complejidad para evitar sobreajuste.

Fortaleza: parsimonia.

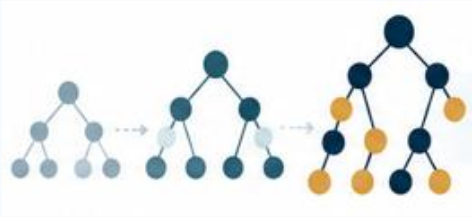


Árboles

Random Forest · XGBoost

Particionan el espacio de predictores y capturan interacciones y umbrales.

Fortaleza: no linealidades.



Árbol Bayesiano

BART

Suma muchos árboles débiles dentro de un marco probabilístico.

Fortaleza: regularización + incertidumbre.

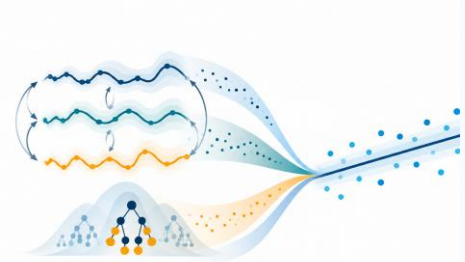


Híbrido

VAR + BART

El VAR captura dinámica lineal; BART corrige los residuos no lineales.

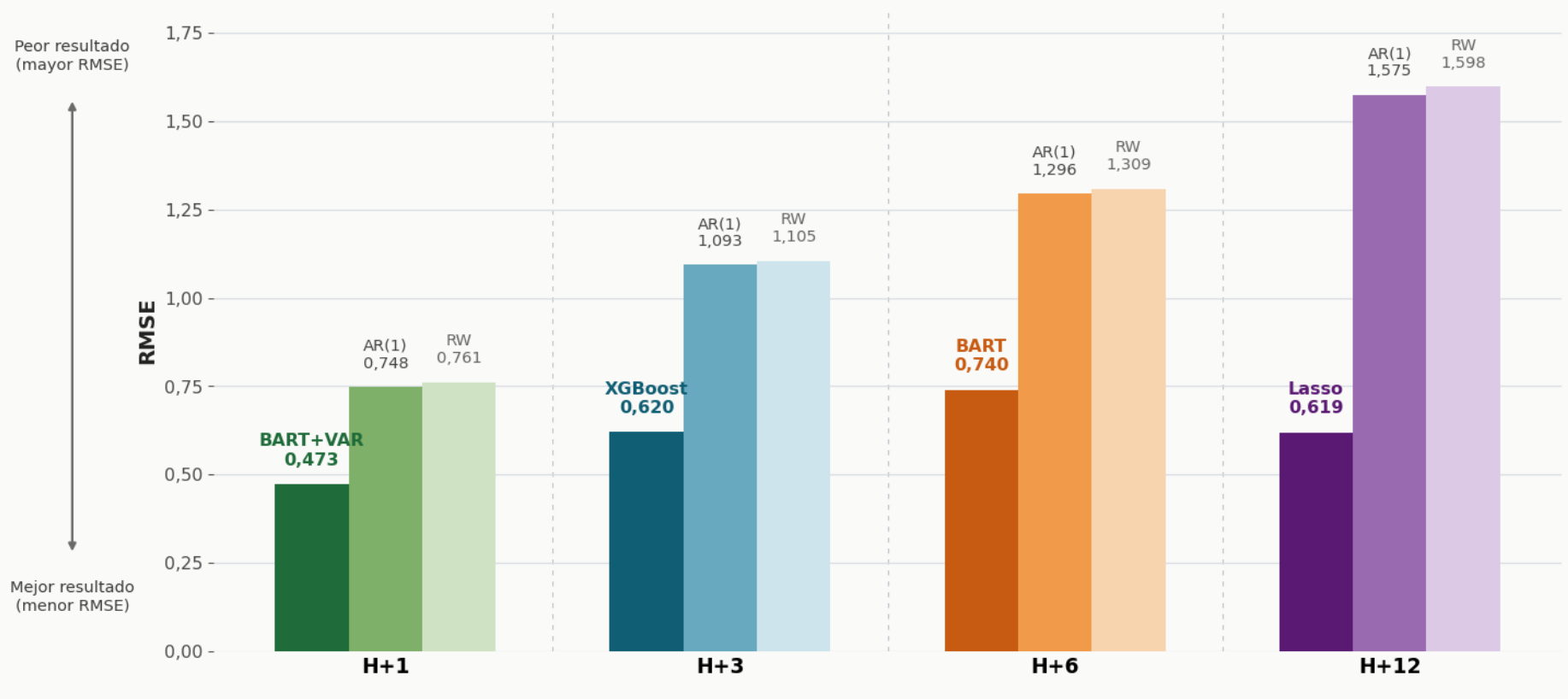
Fortaleza: estructura + flexibilidad.



LA INVESTIGACIÓN

Primer hallazgo: el horizonte importa tanto como el algoritmo.

Raíz del error cuadrático medio del modelo ganador y referencias según horizonte



Nota: cálculos para el periodo enero 2024 a diciembre 2025.
Fuente: Elaboración propia.

BART+VAR
Mejor pronóstico a un mes

XGBoost
Mejor pronóstico a tres meses

BART
Mejor pronóstico a seis meses

Lasso
Mejor pronóstico a doce meses

RMSE: mide qué tan lejos están, en promedio, los pronósticos de los valores observados.

LA INVESTIGACIÓN

¿Qué significa esto desde una perspectiva económica?

Los modelos no lineales funcionan bien porque algunas relaciones económicas no son constantes.

Tipo de cambio → inflación

¿La inflación se comporta de igual modo con una depreciación de 1% que con una de 10%?

La relación entre dos variables puede cambiar.

Salarios → precios internos

¿Un aumento fuerte de los salarios se transmite igual que una caída?

Puede haber asimetrías, rigideces y ajustes discretos.

Expectativas → inflación

¿Importan siempre igual o se vuelven más informativas al alejarse de ciertos niveles?

Pueden existir umbrales entre la relación de dos variables.

Estos modelos no prueban causalidad, ayudan a detectar una estructura predictiva.

LA INVESTIGACIÓN

¿Qué nos dicen los valores de Shapley?

Una forma sencilla de abrir la “caja negra”: repartir la contribución del pronóstico entre las variables.

¿Cómo interpretarlos?

Imagine que el pronóstico es el resultado de un equipo.
Shapley reparte el “crédito” entre sus integrantes.

Pronóstico = base + aportes

+ empuja al alza

– contiene

No dice qué causa la inflación. Dice cómo el modelo llegó a ese pronóstico.

Resultados para modelo XGBoost: la señal cambia con el horizonte

H+1

Persistencia + costos

Inflación actual · salario mínimo real · IPP · expectativas

H+3

Costos + inflación importada

IPP · importaciones · tipo de cambio

H+6

Canal externo + financiero

Comercio exterior · tipo de cambio · TBP / prime

H+12

Política monetaria + fundamentos

La TPM pasa a ser la variable dominante

Mensaje: inercia y costos → canales externos + financieros → política monetaria

La investigación respondió una pregunta, pero surgieron muchas más:

¿Cómo podemos mejorar e integrar el uso de estos nuevos modelos al ejercicio de pronósticos de corto plazo del BCCR?

EXPERIMENTO ORIGINAL

Conjunto de datos “pequeño”

Inflación interanual directa

Horizontes: H+1 · H+3 · H+6 · H+12

Competencia entre modelos

Pronóstico principalmente puntual

Interpretación SHAP solo en XGBoost

SISTEMA ACTUAL

Es posible ampliar el conjunto de datos

Variación mensual del IPC captura mejor las señales más recientes

Enfoque en el corto plazo

Uso de modelos “comparables”

Distribución predictiva

Mayor uso de herramientas interpretabilidad de los resultados

Pronósticos combinados

Escenarios condicionados

Objetivo: Pasar de un ejercicio académico a un sistema de generación de pronósticos de corto plazo.

NUEVA VERSIÓN

Primera mejora: pronosticar el proceso mensual y reconstruir la trayectoria

Seis modelos directos por horizonte, información disponible en tiempo real y coherencia del IPC.

Datos disponibles

Conjunto de datos más amplio (130 variables).

Considera los rezagos de publicación.

Variable de interés

Variación mensual del IPC.

Enfoque en corto plazo

Un modelo directo para cada horizonte:
H+1
…
H+6

IPC

Se reconstruye el nivel del índice.

Interanual

La variación interanual se deriva de la trayectoria.

Ventaja: cada mes obtenemos una trayectoria completa, no cuatro puntos aislados.

NUEVA VERSIÓN

Los modelos: cuatro algoritmos orientados a resolver el mismo problema

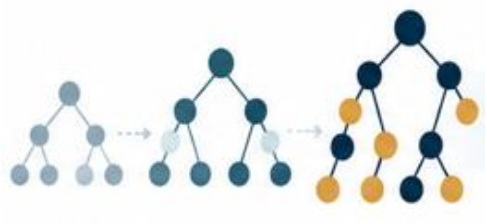
Se seleccionan algoritmos que capturan la no linealidad de formas distintas y que muestran un buen desempeño en el corto plazo.

Gradiente extremo de potenciación (XGBoost)

*Potenciación s*ecuencial

Cada árbol intenta corregir lo que los anteriores dejaron sin explicar.

Muy flexible para interacciones.



Bosques aleatorios (RF)

Bagging

Muchos árboles independientes, cada uno ve muestras y variables diferentes.

Robusto y estable.



Árboles de regresión aditivos bayesianos (BART)

Bayesiano

Suma de árboles pequeños bajo regularización probabilística.

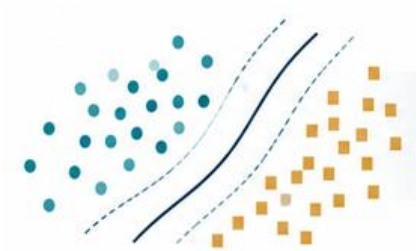
Entrega posterior predictiva.



Máquinas de soporte vectorial (SVM)

Busca una función suave en un espacio transformado de alta dimensión.

Muy sensible a escalas y regularización.

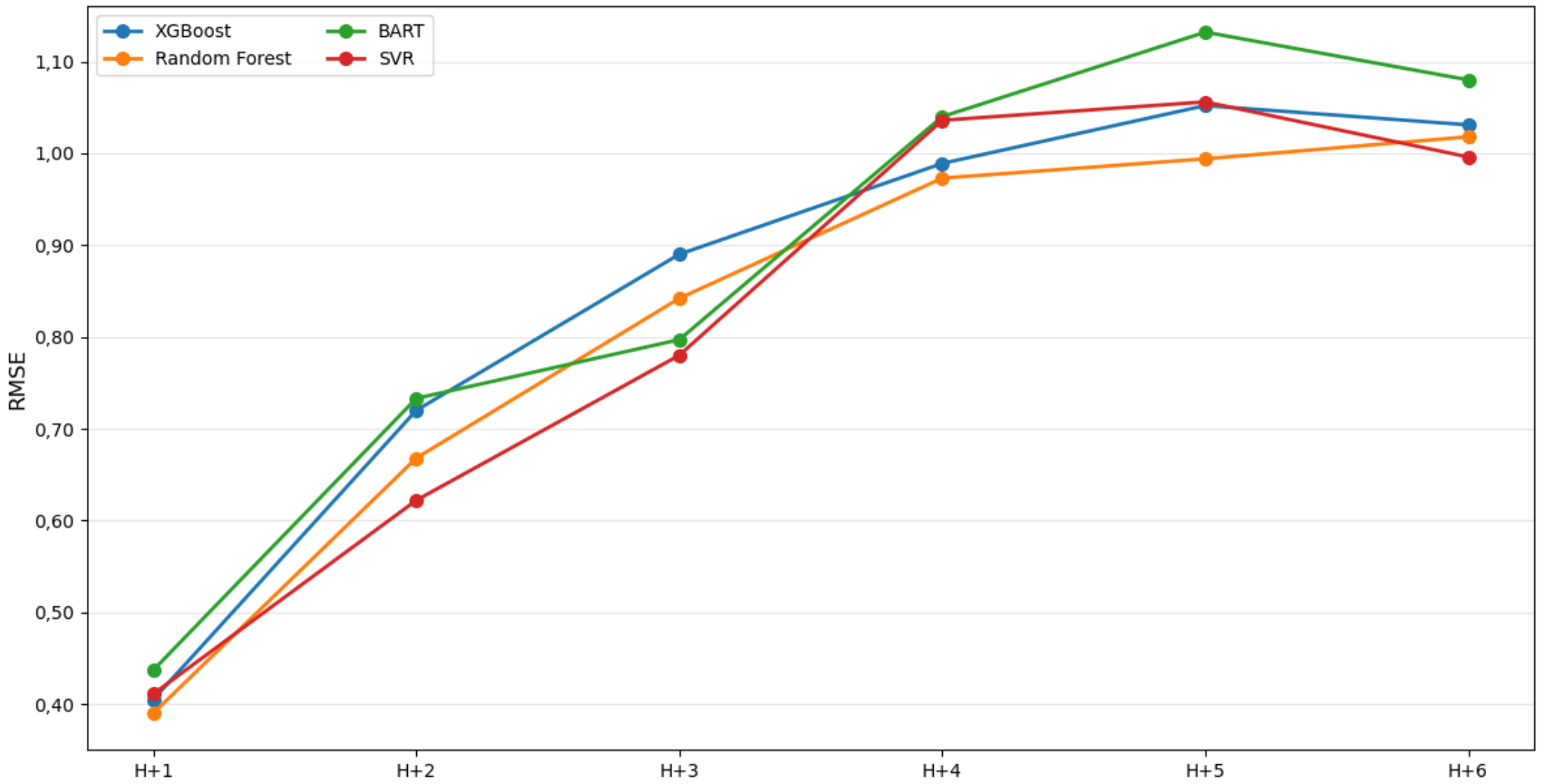


Misma base · misma variable dependiente · mismos rezagos de publicación · misma evaluación

NUEVA VERSIÓN

Resultados congruentes con el ejercicio inicial: No existe un único ganador.

RMSE según modelo según horizonte



H+1	RF
H+2	SVR
H+3	SVR / BART
H+4	RF / XGB
H+5	RF
H+6	SVR

Nota: cálculos para el periodo enero 2024 a setiembre 2026.
Fuente: Elaboración propia.

Ningún modelo domina todos los horizontes.

NUEVA VERSIÓN

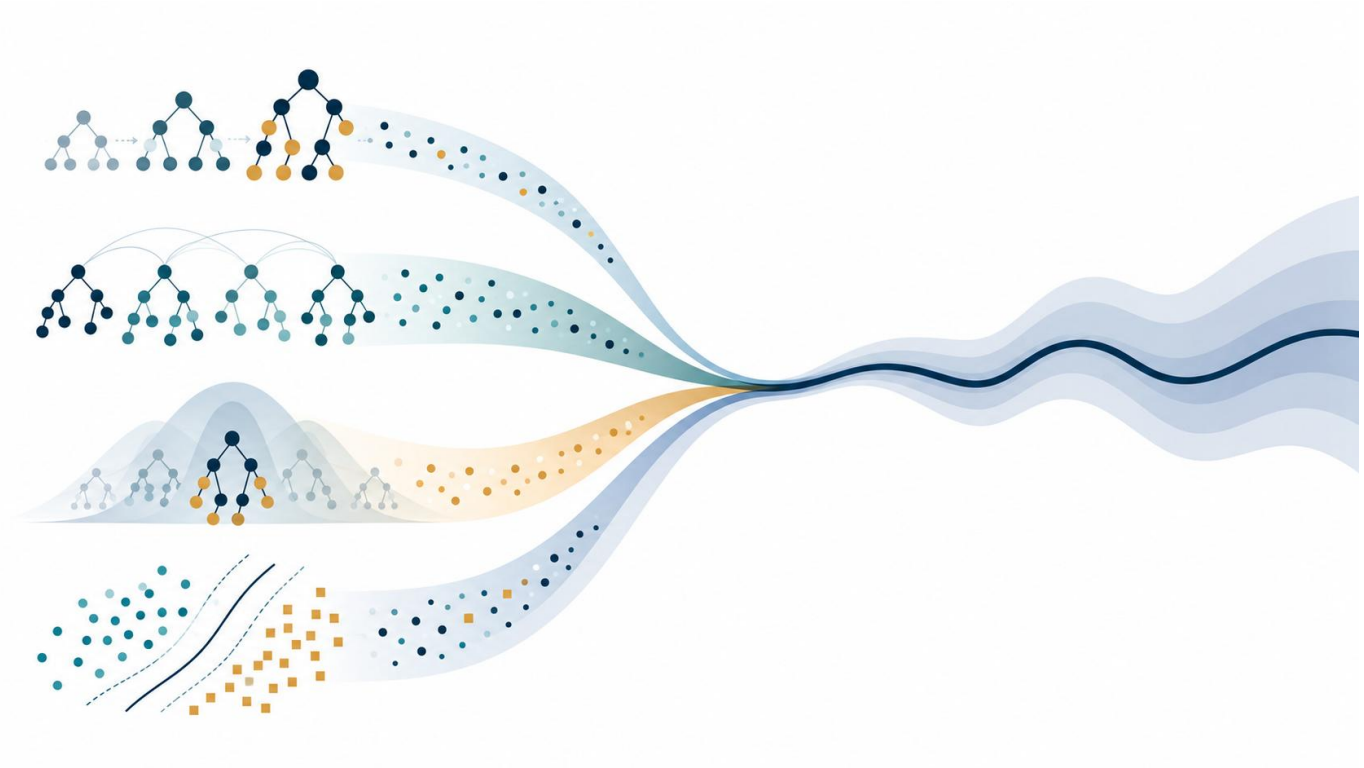
No es necesario elegir un modelo, es posible trabajar con pronósticos combinados

Ventajas principales

Mayor robustez: no se depende de un modelo particular.

Complementariedad: Aprovecha la ventaja de los modelos en distintos horizontes.

Mejor representación de la incertidumbre: permite la combinación de las distribuciones de cada modelo.



Combinación con pesos iguales

25% XGBoost · 25% RF · 25% BART · 25% SVR

Combinación ponderada por RMSE

Mayor peso a los modelos con menor error histórico en cada horizonte.

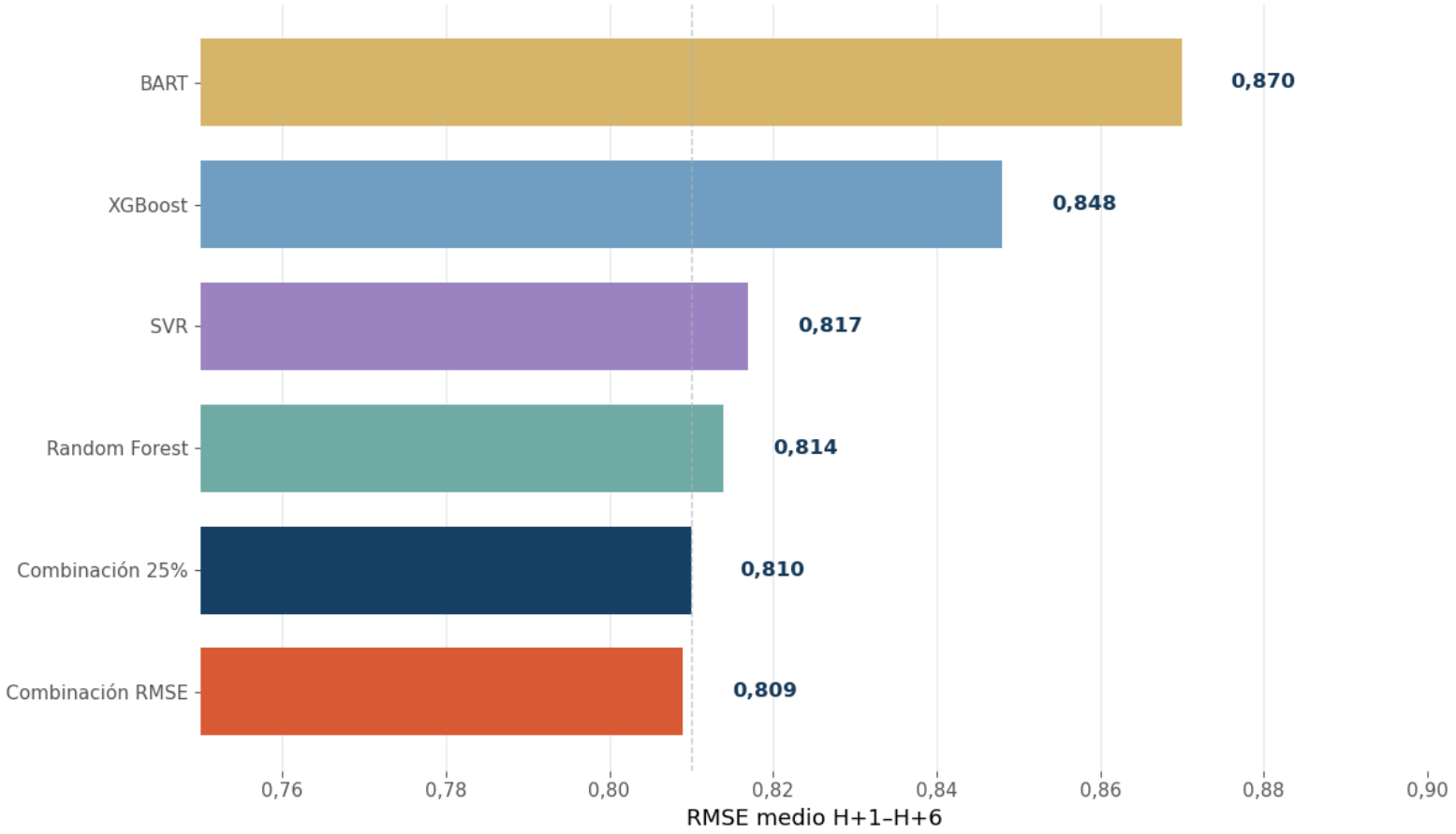
NUEVA VERSIÓN

La ganancia principal de combinar los resultados es la robustez

El uso de pronósticos combinados elimina el riesgo de tener que decidir por algún algoritmo particular.

RMSE según modelos y tipo de combinación

Promedio para todos los horizontes



Mejor RMSE medio

Pronóstico combinado supera a cualquier pronostico mediante modelos individuales.

Modelos individuales líderes

RF ≈ 0.814 · SVR ≈ 0.817
RMSE promedio para todos los horizontes.

¿Ponderar o no?

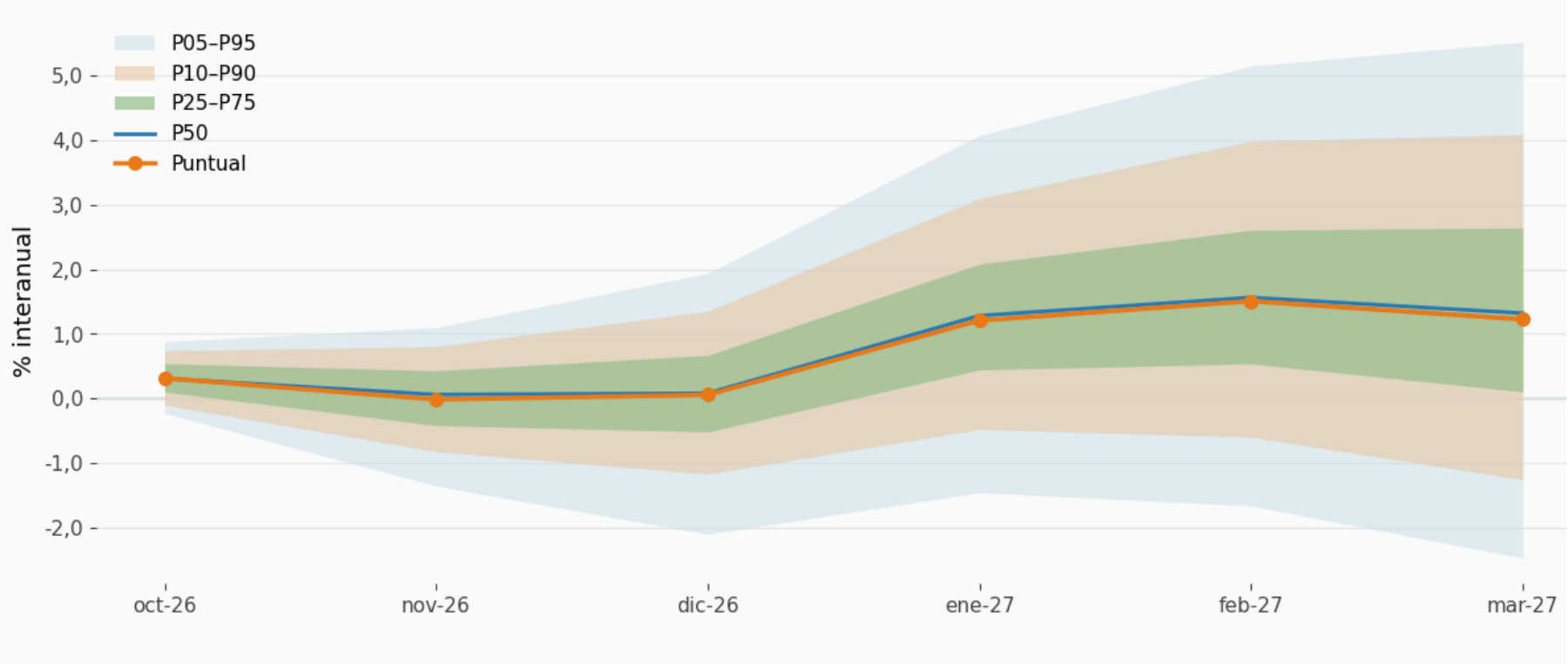
No parece haber mucha diferencia entre los métodos de combinación valorados.

Nota: cálculos para el periodo enero 2024 a setiembre 2026.
Fuente: Elaboración propia.

NUEVA VERSIÓN

La incertidumbre también puede pronosticarse:

Pronóstico de la inflación: trayectoria interanual y distribución predictiva



Nota: para el cálculo de la distribución combinada se promedian los deciles de cada una de las distribuciones individuales.
Fuente: Elaboración propia.

Potenciación
Cuantiles

Bosques aleatorios
Árboles

Árboles de regresión bayesianos
Posterior

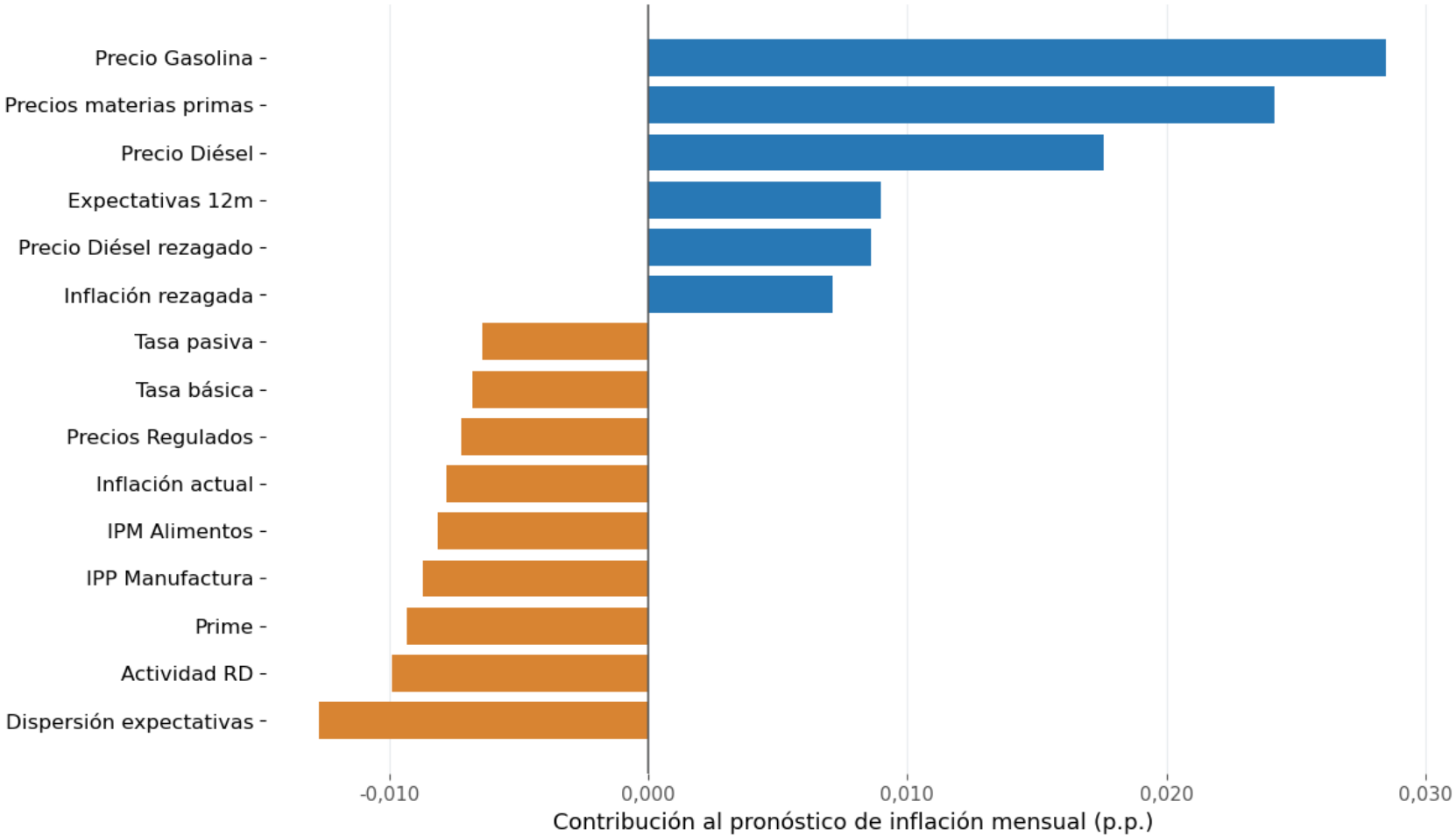
Máquinas de soporte vectorial
Bootstrap

NUEVA VERSIÓN

También podemos preguntar: “¿qué explica el valor del pronóstico?”

SHAP permite descomponer un pronóstico puntual en contribuciones de las variables.

Valores SHAP del pronóstico mensual del modelo Random Forest para octubre 2026



Empuja al alza

Gasolina súper · materias primas · diésel

Contiene

Expectativas · IMAE · tasas y precios rezagados

Interpretación del modelo ≠ causalidad económica

Fuente: Elaboración propia.

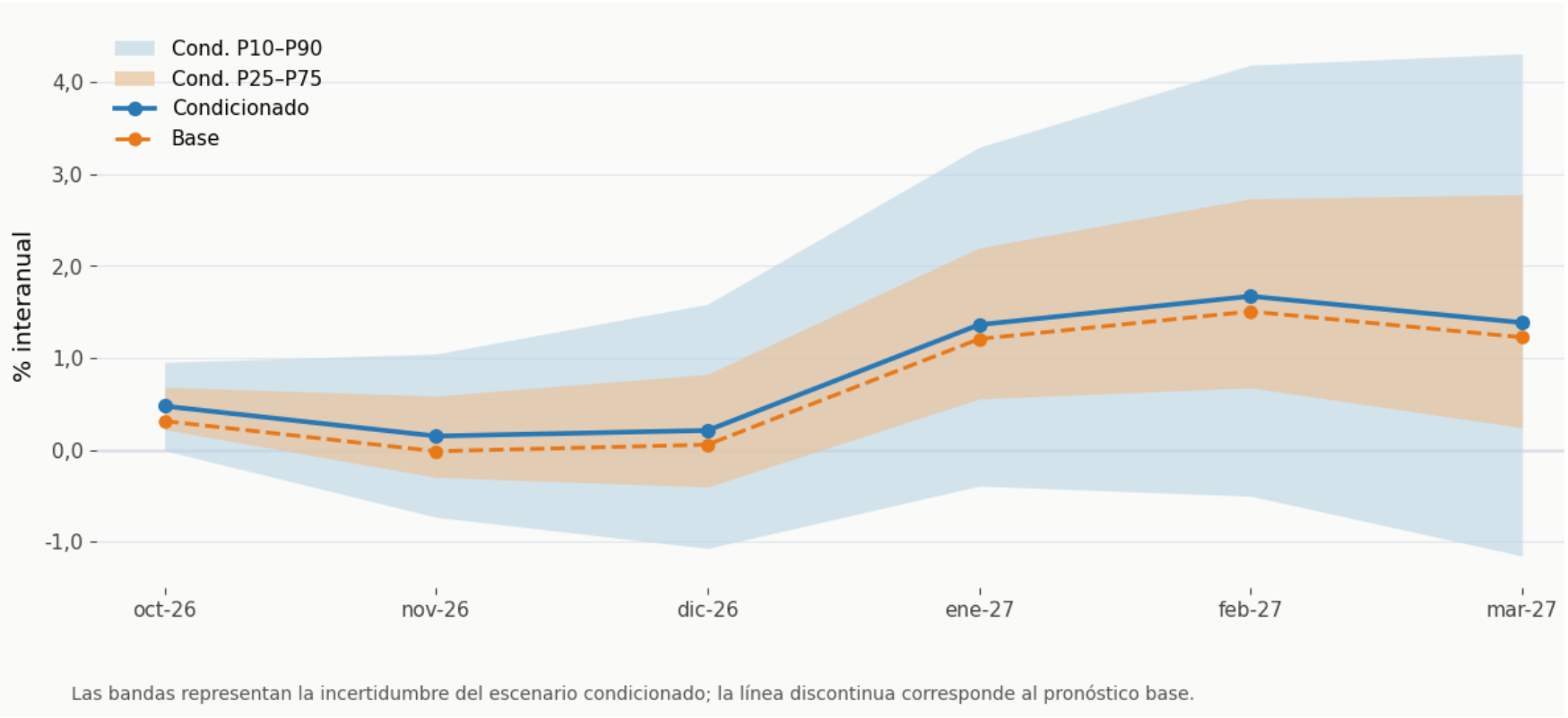
NUEVA VERSIÓN

Y finalmente podemos preguntar: “¿qué pasa si…?”

Los modelos pueden incorporar supuestos explícitos sobre algunas variables y comparar una trayectoria base con una condicionada.

- **Combustibles (precios locales)**
- **Precio del petróleo e índice de materias primas**
- **Tipo de cambio nominal**
- **Salarios mínimos**
- **Expectativas de inflación**

Pronóstico combinado ponderado por RMSE: Base y condicionado



Pronóstico condicionado: Incremento hipotético de 10% en el precio del diésel, la gasolina súper y regular en H+1.
Fuente: Elaboración propia.

Lecciones aprendidas

01 Los modelos de aprendizaje automático agregan valor principalmente en el corto plazo

Pueden capturar no linealidades, interacciones y gran cantidad de información simultáneamente.

02 No existe un modelo universalmente mejor

El horizonte, y probablemente el estado de la economía, definen qué estructura funciona mejor.

03 El mayor valor está en generar un proceso de formación de pronósticos

Pronóstico + incertidumbre + combinación de modelos + explicación + escenarios.

BANCO CENTRAL DE COSTA RICA

JORNADAS DE INVESTIGACIÓN ECONÓMICA 2026

Métodos de aprendizaje automático para la predicción de la inflación en múltiples horizontes: **un análisis para Costa Rica**

Inv_Eco@bccr.fi.cr

Autor: Manuel Esteban Sánchez Gómez

17 de noviembre, 2026