

ARTÍCULO DE REVISIÓN BIBLIOGRÁFICA

EVALUACIÓN DE LA PRECISIÓN EN EL PRONÓSTICO DE LA INFLACIÓN EN BOLIVIA: RANDOM FOREST Y ÁRBOLES DE DECISIÓN VS. ARIMA

EVALUATION OF FORECAST ACCURACY FOR INFLATION IN BOLIVIA: RANDOM FOREST AND DECISION TREES VS. ARIMA



José Antonio Zurita Herrera

Profesor Universidad Privada del Valle, Bolivia.

jzuritah@univalle.edu

Compás
empresarial

Recibido: 17/11/2024 Revisado: 11/12/2024 Aceptado: 12/12/2024

Citar: ZURITA HERRERA, J. A. IF EVALUACIÓN DE LA PRECISIÓN EN EL PRONÓSTICO DE LA INFLACIÓN EN BOLIVIA: RANDOM FOREST Y ÁRBOLES DE DECISIÓN VS. ARIMA. *Revista Compás Empresarial*, 15(39). <https://doi.org/10.52428/20758960.v15i39.1227>

Nota: Los autores declaran no tener conflicto de intereses con respecto a esta publicación y se responsabilizan de contenido vertido.

Fuentes de financiamiento: Esta investigación fue financiada con fondos de los autores.

RESUMEN

Este estudio compara modelos de pronóstico del IPC en Bolivia, evaluando enfoques tradicionales y de aprendizaje automático para la predicción de inflación. Se aplicaron modelos de Árbol de Decisión, Árbol Podado, Random Forest y ARIMA, encontrando que los modelos de aprendizaje automático, especialmente el Árbol Podado, superan al ARIMA en precisión, con un menor RMSE en el conjunto de testeo. Esto sugiere que los modelos modernos capturan mejor las dinámicas complejas del IPC y representan herramientas más robustas para la proyección de inflación en economías emergentes. El estudio recomienda explorar modelos híbridos y redes neuronales avanzadas en futuras investigaciones para optimizar aún más los pronósticos.

Palabras clave: Pronóstico de inflación, aprendizaje automático, Random Forest, Árbol de Decisión, modelo ARIMA, Bolivia



Esta obra está bajo una licencia internacional [Creative Commons Atribución 4.0](https://creativecommons.org/licenses/by/4.0/).
Derechos de autor 2024, José Antonio Zurita Herrera

ABSTRACT

This study compares IPC forecasting models in Bolivia, evaluating traditional and machine learning approaches for inflation prediction. Decision Tree, Pruned Tree, Random Forest, and ARIMA models were applied, finding that machine learning models, particularly the Pruned Tree, outperform ARIMA in accuracy, achieving a lower RMSE in the test set. This suggests that modern models better capture the complex dynamics of the IPC and represent more robust tools for inflation projection in emerging economies. The study recommends exploring hybrid models and advanced neural networks in future research to further optimize forecasts.

Keywords: Inflation Forecasting, Machine Learning, Random Forest, Decision Tree, ARIMA Model, Bolivia.

INTRODUCCIÓN

En el contexto actual de Bolivia, el pronóstico de inflación ha cobrado gran relevancia debido al reciente incremento en los precios y el temor a una posible hiperinflación similar a la de los años 80. Durante esa época, el país sufrió una de las peores crisis inflacionarias de América Latina, que resultó en una pérdida significativa del poder adquisitivo y desconfianza en las políticas públicas. Aunque el contexto económico ha cambiado, las presiones inflacionarias actuales, tanto internas como externas, evidencian la necesidad de contar con herramientas adecuadas para anticipar y gestionar estas variaciones en el Índice de Precios al Consumidor (IPC).

La inflación es una variable compleja influenciada por política monetaria, precios de materias primas y condiciones internacionales, lo que dificulta el uso de métodos tradicionales como ARIMA, que tienen limitaciones para capturar relaciones no lineales en economías complejas. Los modelos de aprendizaje automático, como el Random Forest, superan estas limitaciones al manejar alta dimensionalidad y detectar relaciones no lineales, adaptándose mejor a la volatilidad de factores económicos en países emergentes.

Este estudio compara la precisión de modelos de árboles de decisión (árbol simple, podado y Random Forest) con el modelo ARIMA para pronosticar la inflación en Bolivia, utilizando datos mensuales de variables económicas clave. La evaluación se basa en el RMSE para determinar el modelo más preciso en este contexto. Se busca ofrecer herramientas predictivas robustas y adaptables que apoyen la formulación de políticas económicas y contribuyan a la estabilidad de precios y el bienestar a largo plazo en Bolivia.

REVISIÓN BIBLIOGRÁFICA

El pronóstico de la inflación ha sido una preocupación constante en economía debido a su impacto en la política monetaria y económica. Tradicionalmente, los modelos ARIMA han predominado por su simplicidad y capacidad para analizar series temporales, y el enfoque de Box y Jenkins (1970) ha sido ampliamente aplicado en este campo. En países con alta volatilidad inflacionaria, como Bolivia, estos modelos han sido útiles para captar patrones estacionales y de corto plazo (Bojanic, 2021). Sin embargo, la evolución de la tecnología y la disponibilidad de datos han impulsado el uso de métodos de aprendizaje automático en el pronóstico inflacionario.

Modelos basados en árboles, como los árboles de decisión y Random Forest, han demostrado ser más precisos y adaptables en contextos económicos volátiles al superar los enfoques tradicionales en predicciones inflacionarias, gracias a su capacidad para manejar datos de alta dimensionalidad y capturar relaciones no lineales (Medeiros et al., 2021). Introducido por Breiman (2001), el método de Random Forest destaca por su habilidad para reducir el sobreajuste y mejorar la precisión en los pronósticos.

Estudios como el de Forte (2024) han demostrado la eficacia de Random Forest en contextos inflacionarios complejos, destacando su precisión superior a modelos tradicionales como ARMA y comparable a técnicas avanzadas como Ridge y Lasso, con un menor Error Absoluto Medio. Por otro lado, Zhao y Zhang (2019) subrayan la efectividad de redes neuronales, como RNN y LSTM, para capturar variaciones de corto plazo y responder a eventos inesperados, especialmente en entornos de alta incertidumbre.

El uso de datos no convencionales, como las tendencias de búsqueda en Google, es útil para captar percepciones públicas sobre la inflación en momentos de alta incertidumbre (Eugster y Uhl, 2024). En África, los modelos de árboles de decisión y Random Forest mejoran la precisión de los pronósticos inflacionarios (Moyo y Musengezi, 2022). La combinación de modelos autorregresivos con algoritmos de árboles ofrece predicciones robustas en entornos volátiles (Patton y Simsek, 2023). Random Forest es adecuado para pronósticos a largo plazo (Meuller, 2022) y la poda de estos modelos mejora la precisión (Zhou y Mentch, 2023). El Banco Central Europeo (2023) encontró que el Quantile Regression Forest es eficaz para capturar patrones no lineales en la Zona Euro.

La literatura reciente destaca que los métodos basados en árboles ofrecen ventajas significativas frente a los modelos tradicionales, especialmente en economías

emergentes. Altansukh et al. (2017) encontraron que los modelos de aprendizaje automático son altamente efectivos en América Latina para manejar entornos económicos volátiles. En Brasil, estudios como los de Garcia et al. (2017) evidencian que Random Forest supera a los modelos AR y de paseo aleatorio en pronósticos inflacionarios, adaptándose mejor a la complejidad económica. Asimismo, Silva y Oliveira (2019) demostraron que Random Forest produce predicciones más precisas en comparación con métodos convencionales en el mismo país.

En México, Espinosa y Zúñiga (2020) encontraron que algoritmos como Random Forest y XGBoost mejoran la precisión de los pronósticos inflacionarios. En Colombia, Carmona y López (2021) destacaron el uso de Random Forest para analizar grandes conjuntos de datos. En Perú, Medina y Chacón (2017) resaltaron su utilidad en el análisis de múltiples variables para predecir la inflación. En resumen, los métodos de aprendizaje automático basados en árboles, como Random Forest, superan a los modelos tradicionales en precisión, especialmente en economías emergentes, al capturar relaciones no lineales y fluctuaciones complejas.

METODOLOGÍA

Esta sección describe metodologías supervisadas, como árboles de decisión, árboles podados, Random Forest y modelos ARIMA, destacando su utilidad en problemas de clasificación y regresión.

Árbol de Decisión Simple

El árbol de decisión organiza los datos en una estructura jerárquica donde cada nodo representa una decisión sobre una variable, y cada rama muestra el resultado de esa decisión. Para lograr particiones homogéneas, el modelo utiliza criterios como la ganancia de información y el índice Gini, que seleccionan las variables que mejor maximizan la homogeneidad en cada partición. La ganancia de información se calcula a partir de la entropía del conjunto de datos:

$$H(S) = - \sum_{i=1}^n p_i \log_2 (p_i) \quad (1)$$

donde $H(S)$ representa la entropía del conjunto S con n categorías, y p_i es la proporción de observaciones en la categoría i (Quinlan, 1986). A partir de esta entropía, la ganancia

de información al dividir el conjunto S en función de una característica A se calcula mediante:

$$Ganancia(S, A) = H(S) - \sum_{v \in \text{valores de } A} \frac{|S_v|}{|S|} H(S_v) \quad (2)$$

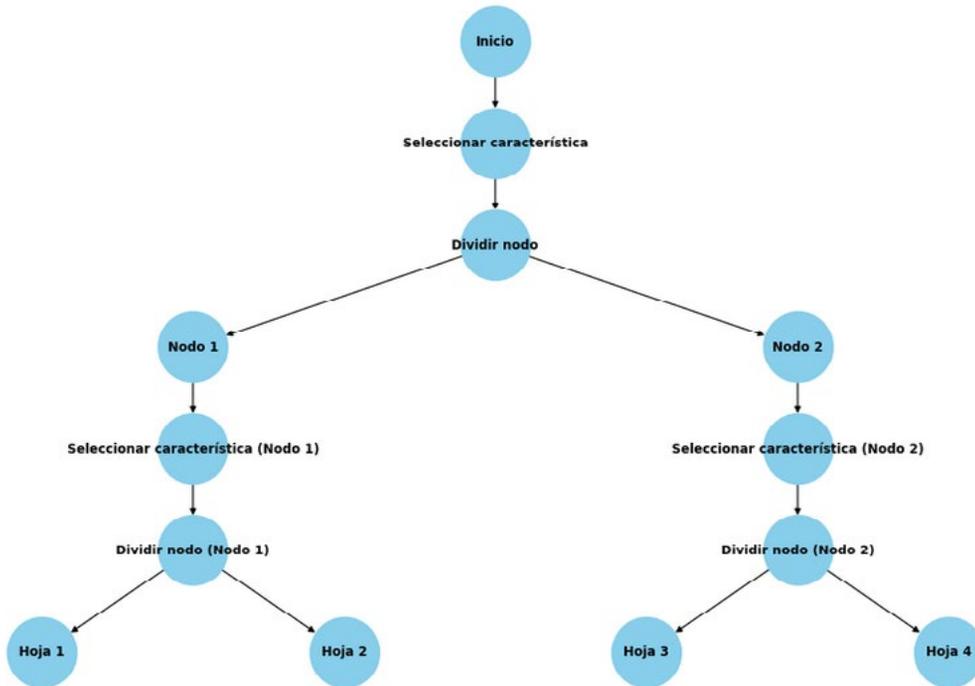
Para este caso, el índice Gini mide la impureza en la distribución de las clases en S y se calcula como:

$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (3)$$

donde p_i representa la proporción de cada clase en el conjunto (Breiman et al., 1984). Estos criterios permiten al modelo seleccionar la partición más adecuada en cada nodo del árbol de decisión.

La ilustración 1 muestra los pasos básicos para construir un árbol de decisión: inicia con los datos de entrenamiento, selecciona la característica óptima para dividirlos, divide el nodo en dos ramas, y repite el proceso en cada nodo resultante. Los nodos finales, o hojas, representan las predicciones finales. El proceso continúa hasta cumplir con los criterios de parada, como la profundidad máxima o la pureza de los nodos.

Ilustración 1. Construcción de un árbol de decisión



Fuente: Elaboración propia, 2024

Árbol de Decisión Podado

Para mejorar la capacidad de generalización de los árboles de decisión, se emplea el podado de árboles, una técnica que elimina las ramas que aportan poco a la predicción. Entre los métodos de podado más comunes están el criterio de complejidad del costo y el podado por validación cruzada. El criterio de complejidad del costo combina el error de predicción y el número de hojas en el árbol para evitar el sobreajuste. Se calcula el costo total de un árbol T de la siguiente manera:

$$\text{Costo}(T) = \text{Error}(T) + \alpha \times |T| \quad (4)$$

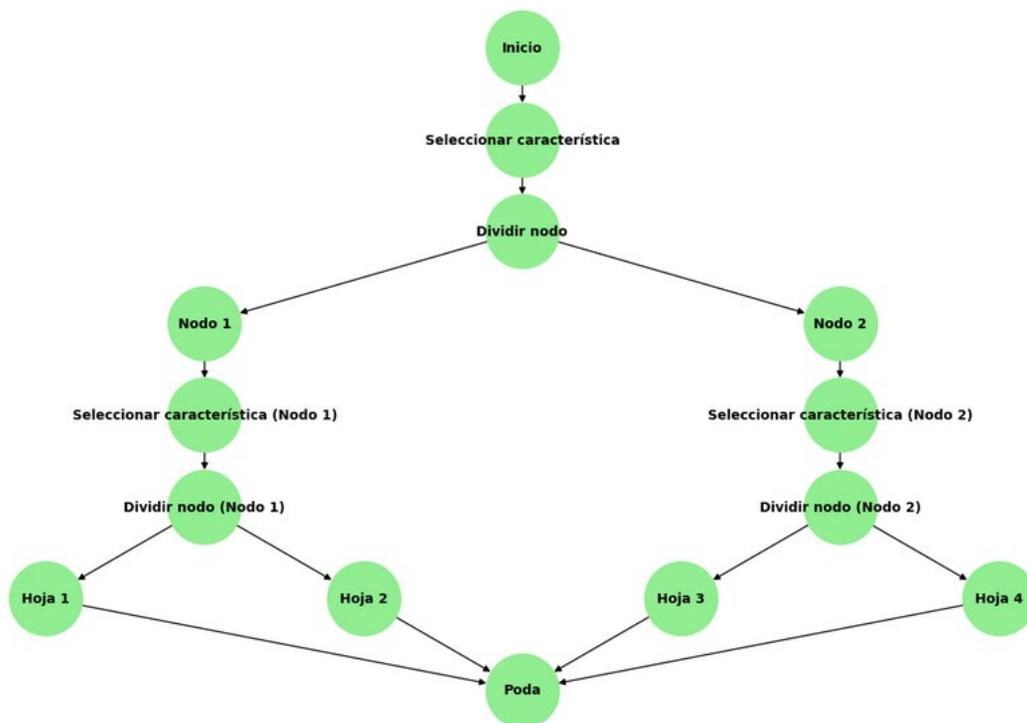
donde $\text{Error}(T)$ es el error de predicción del árbol T ; $|T|$ es el número de hojas del árbol, y α es un parámetro de regularización que penaliza la complejidad (Hastie *et al.*, 2009).

Esta fórmula selecciona el tamaño del árbol que minimiza el costo total, evitando la complejidad innecesaria.

El criterio de podado por validación cruzada utiliza un conjunto de validación para evaluar el error en diferentes profundidades del árbol. El modelo se poda hasta alcanzar la profundidad que minimiza el error en el conjunto de validación, lo cual ayuda a evitar el sobreajuste sin necesidad de complejidad adicional en el cálculo (Breiman et al., 1984).

La ilustración 2 detalla los pasos para construir y podar un árbol de decisión: iniciar con los datos de entrenamiento, seleccionar la mejor característica para dividirlos, generar ramas y hojas con predicciones finales, y realizar una poda para eliminar ramas poco relevantes, reduciendo el sobreajuste y mejorando la generalización del modelo.

Ilustración 2. Construcción de un Árbol de Decisión Podado



Fuente: Elaboración propia, 2024

Random Forest

El algoritmo Random Forest combina múltiples árboles de decisión entrenados en muestras aleatorias (bagging) para mejorar la precisión y reducir la varianza en los pronósticos. Las predicciones de los árboles se agregan, lo que aumenta la estabilidad del modelo.

En Random Forest, cada árbol se entrena en una muestra aleatoria S_b , seleccionada con reemplazo del conjunto de datos original. Cada muestra se representa como:

$$S_b = \{(x_i, y_i) | i \in \text{muestra aleatoria con reemplazo}\} \quad (5)$$

donde x_i y y_i son los datos y etiquetas de cada observación en la muestra. En cada nodo, el árbol selecciona solo un subconjunto aleatorio de características para reducir la correlación entre los árboles (Breiman, 2001).

Una vez entrenados todos los árboles, el modelo Random Forest realiza una predicción final combinando los resultados individuales. En problemas de clasificación, como predecir niveles de inflación (baja, moderada o alta), el modelo utiliza la votación mayoritaria de los árboles:

$$\hat{y} = \text{modo}(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_B) \quad (6)$$

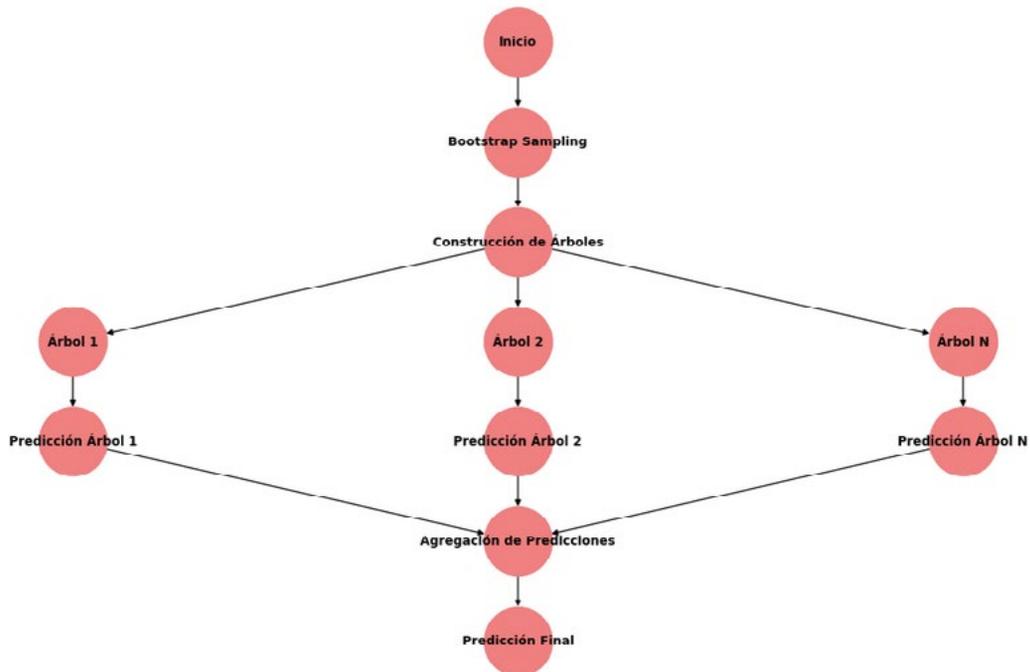
En problemas de regresión, donde el objetivo es un valor continuo, la predicción final es el promedio de todas las predicciones individuales:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}_b \quad (7)$$

donde B es el número de árboles en el bosque y \hat{y}_b representa la predicción del árbol b (Breiman, 2001). Este proceso reduce la varianza del modelo al combinar las predicciones de múltiples árboles.

La ilustración 3 describe los pasos básicos para construir un modelo Random Forest: iniciar con los datos de entrenamiento, aplicar bootstrap sampling para generar muestras aleatorias, construir múltiples árboles de decisión con diferentes subconjuntos de datos y características, realizar predicciones individuales de cada árbol, y combinar estas predicciones mediante votación mayoritaria (clasificación) o promedio (regresión) para obtener la predicción final.

Ilustración 3. Construcción de un Random Forest



Fuente: Elaboración propia, 2024.

Raíz del Error Cuadrático Medio

La Raíz del Error Cuadrático Medio (RMSE) es una métrica utilizada para evaluar la precisión de los modelos de predicción en regresión. Mide la diferencia entre los valores predichos y los observados, indicando la magnitud del error promedio en las mismas unidades de la variable objetivo. Se calcula como la raíz cuadrada de la media de los errores cuadrados entre los valores predichos y observados en un conjunto de datos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

donde:

- y^i : Valor observado en la posición i
- \hat{y}_i : Valor predicho en la posición i
- n : Número total de observaciones

La RMSE proporciona la magnitud promedio del error en las mismas unidades que la variable de interés. Un valor más bajo de RMSE indica un ajuste más preciso del modelo, mientras que un valor alto sugiere un error mayor en las predicciones (Hyndman y Athanasopoulos, 2018). La RMSE es particularmente útil cuando se compara el rendimiento de varios modelos, ya que permite observar cuál tiene un error promedio menor al hacer predicciones.

En este estudio, la variable que se pretende predecir es la inflación a corto plazo, π_t , evaluando la posible relación no lineal con sus valores anteriores y un conjunto de regresores, X_{t-h} , que pueden ser actuales o pasados:

$$\pi_t = f(\pi_{t-1} \dots \pi_{t-h}; x_t \dots x_{t-h}) + e_t \quad (9)$$

Donde e_t representa un término de error y $f(_)$ es una función no lineal sin una forma paramétrica predefinida. A partir de esta estimación, se genera un pronóstico para π_t :

$$\pi_t = f(\pi_{t-1} \dots \pi_{t-h}; x_t \dots x_{t-h}) \quad (10)$$

Es importante aclarar desde el principio que el objetivo de estas estimaciones no es identificar relaciones causales entre las variables, sino centrarse exclusivamente en la predicción. Por lo tanto, los problemas de simultaneidad entre las variables no representan un impedimento para su inclusión en el modelo.

Datos

La base de datos del estudio comprende doce indicadores económicos mensuales (2007-2024), con la inflación mensual como variable principal. Las variables predictoras se seleccionaron siguiendo la teoría de la curva de Phillips e incluyen condiciones monetarias, actividad económica, mercado cambiario y precios internacionales de commodities como petróleo, gas, oro y trigo.

Tabla 1. Variables de estudio

Variable	Indicador	Unidad de medida	Periodo	Frecuencia	Fuente
Inflación (target)	IPC (Índice de Precios al Consumidor)	Variación mensual del IPC en comparación con el mismo mes del año anterior.	Enero 2007 - marzo 2024	Mensual	Instituto Nacional de Estadística - Bolivia
Actividad económica	IGAE (Índice Global de la Actividad Económica)	Variación porcentual del IGAE de un mes en comparación con el mismo mes del año anterior.	Enero 2007 - marzo 2024	Mensual	Instituto Nacional de Estadística - Bolivia
Política monetaria	Tasa de interés activa (efectiva)	Promedios mensuales	Enero 2007 - marzo 2024	Mensual	Banco Central de Bolivia
	Base monetaria	En miles de bolivianos	Enero 2007 - marzo 2024	Mensual	Banco Central de Bolivia
	Agregado monetario M2	En miles de bolivianos	Enero 2007 - marzo 2024	Mensual	Banco Central de Bolivia
	Encaje legal	En porcentaje	Enero 2007 - marzo 2024	Mensual	Banco Central de Bolivia
Precios internacionales	Precio del petróleo	West Texas Intermediate \$us/Barril	Enero 2007 - marzo 2024	Mensual	Banco Mundial
	Precio del trigo	United States (US Gulf Port) \$us/Ton. Métrica	Enero 2007 - marzo 2024	Mensual	Banco Mundial
	Precio del gas	United States \$us/MM BTU (British Thermal Unit)	Enero 2007 - marzo 2024	Mensual	Banco Mundial
	Precio del oro	United Kingdom (London) \$us/Onza Troy	Enero 2007 - marzo 2024	Mensual	Banco Mundial
Mercado cambiario	Tipo de cambio oficial	Bs/\$us	Enero 2007 - marzo 2024	Mensual	Banco Mundial
	Reservas Internacionales Netas	En miles de bolivianos	Enero 2007 - marzo 2024	Mensual	Banco Central de Bolivia

Fuente: Elaboración propia, 2024.

Para la estimación de los modelos de árbol de decisión, árbol podado y random forest, se utilizó el lenguaje de programación Python, aprovechando bibliotecas como scikit-learn y pandas para el procesamiento y análisis de datos. Por otro lado, para la estimación del modelo ARIMA, se empleó el software RStudio, utilizando el lenguaje de programación R y paquetes como forecast y tseries para el análisis de series temporales.

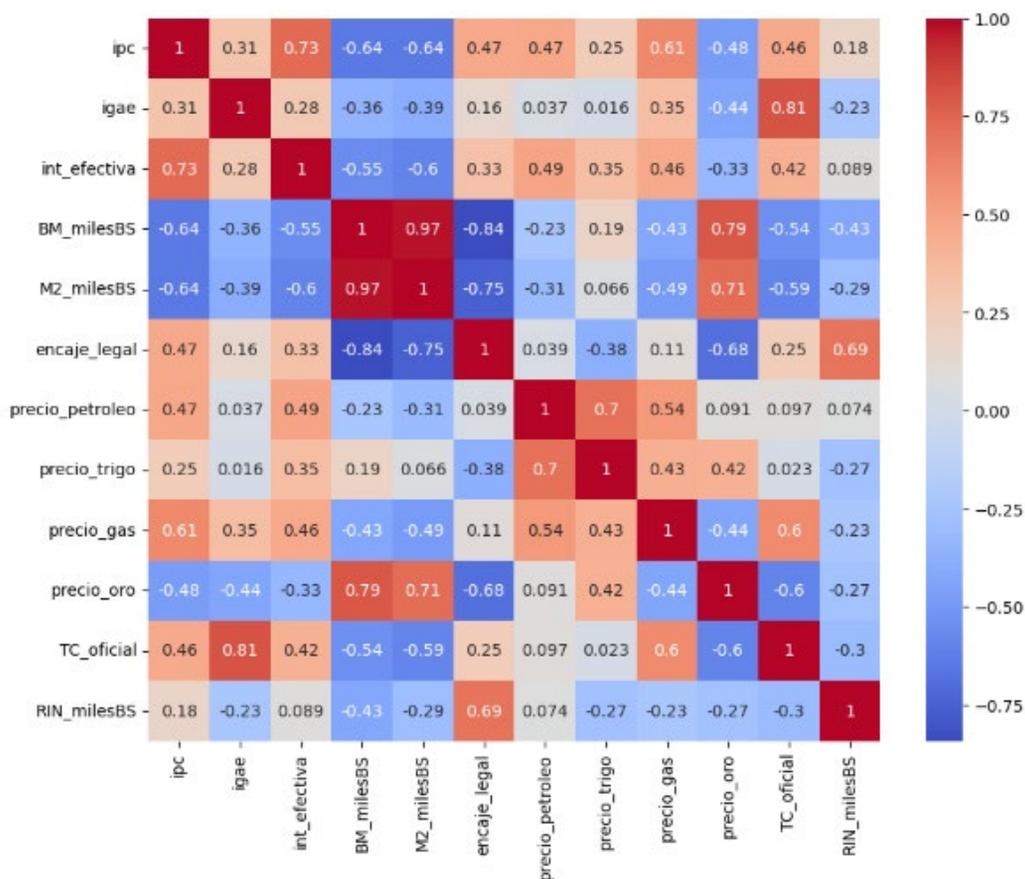
RESULTADOS

En esta investigación se analizaron los modelos de Árbol de Decisión, Árbol Podado y Random Forest, técnicas comunes en el aprendizaje automático, para evaluar su desempeño en el pronóstico a corto plazo de la inflación en Bolivia. Se concluyó sobre su aplicabilidad y el valor añadido que pueden ofrecer como herramientas complementarias para predecir la evolución de los precios en un entorno complejo. Es importante aclarar que estas estimaciones no implican causalidad, sino solo evalúan el rendimiento de los modelos con fines de pronóstico.

Matriz de correlación

La matriz de correlación revela que el IPC en Bolivia tiene una fuerte correlación positiva con la tasa de interés efectiva ($r=0.73$), lo que sugiere que mayores costos de financiamiento podrían trasladarse a los precios finales. Por otro lado, la oferta monetaria M2 muestra una correlación negativa significativa con el IPC ($r=-0.64$), posiblemente debido a políticas monetarias y cambiarias que mitigan el impacto inflacionario de una expansión monetaria mediante el control del tipo de cambio y precios en sectores clave.

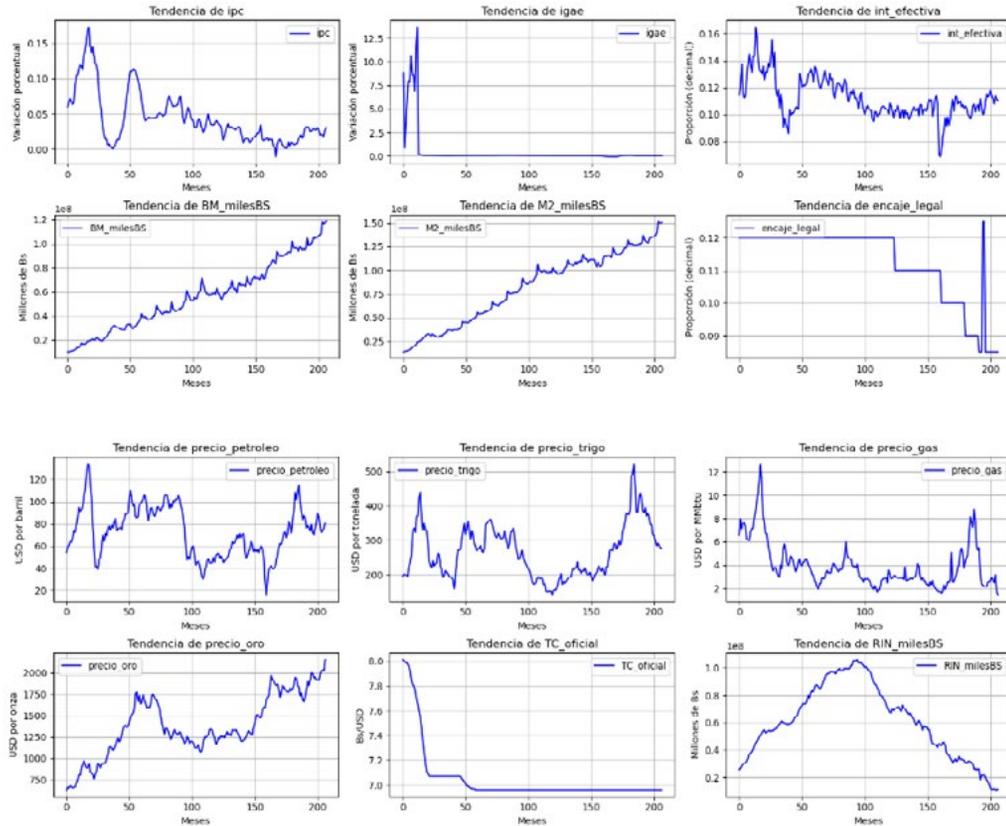
El análisis de la matriz de correlación en Bolivia revela que el IPC está influenciado por factores tanto internos como externos. El encaje legal y el tipo de cambio oficial muestran correlaciones positivas moderadas con el IPC ($r=0.47$ y $r=0.46$), indicando que políticas restrictivas de liquidez y ajustes cambiarios pueden generar presiones inflacionarias. Los precios internacionales de materias primas, como el petróleo ($r=0.47$) y el gas ($r=0.35$), también impactan la inflación, especialmente en sectores de transporte y producción. Las reservas internacionales netas, con una baja correlación ($r=0.18$), tienen un impacto indirecto limitado en el IPC, ayudando a mantener la estabilidad económica y el tipo de cambio.

Gráfico 1. Mapa de calor de correlación de variables

Fuente: Elaboración propia, 2024

En resumen, la inflación en Bolivia está fuertemente influenciada por factores internos como las políticas monetarias y cambiarias, así como por factores externos como los precios internacionales del petróleo. La relación entre el IPC y la oferta monetaria M2 refleja el impacto de políticas de control de precios y la estabilidad cambiaria en el país, lo que puede contrarrestar las presiones inflacionarias. Este análisis proporciona una base para construir modelos predictivos y formular políticas económicas que busquen estabilizar los precios en Bolivia.

Gráfico 2. Tendencia de los Factores Inflacionarios en Bolivia

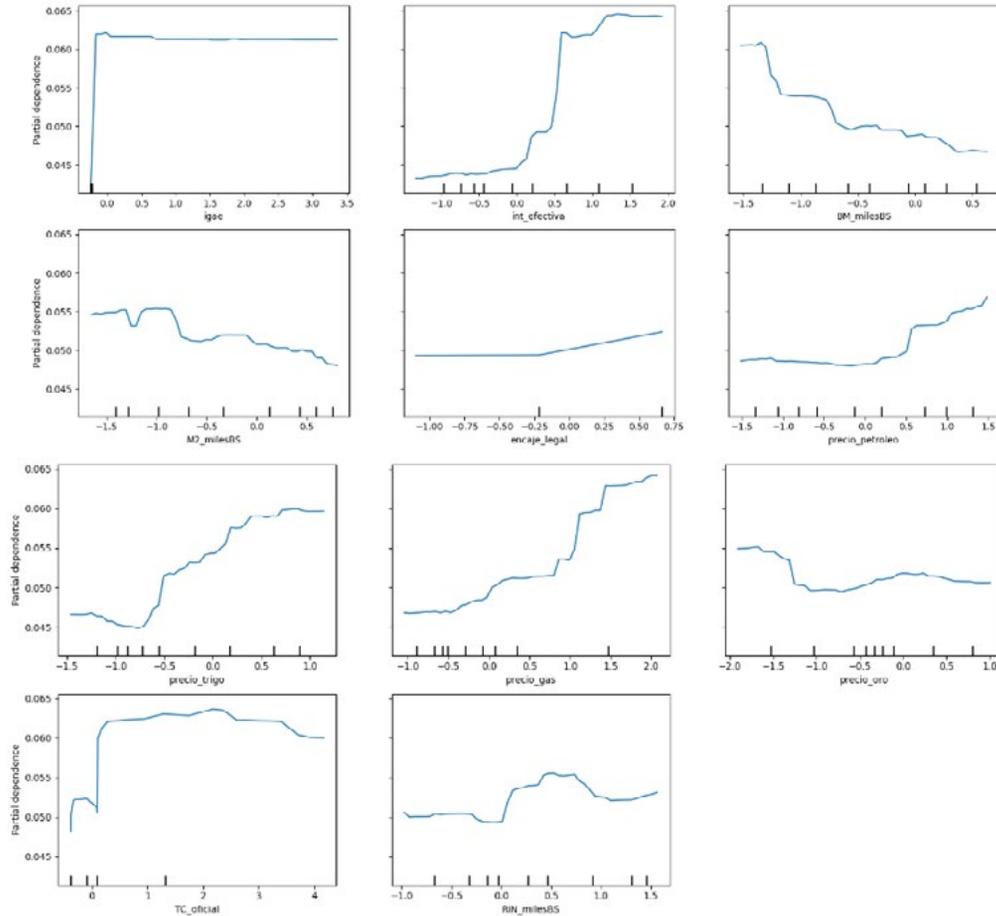


Fuente: Elaboración propia, 2024

Dependencia parcial de la inflación

Los gráficos de dependencia parcial (gráfico 3) muestran el efecto marginal de cada variable independiente (regresor) sobre la inflación mensual (IPC) mientras se mantienen constantes los demás regresores. Estos gráficos permiten identificar cómo el IPC responde a cambios en cada regresor en el contexto del modelo de predicción, proporcionando una interpretación sobre la relación específica de cada variable con la inflación en Bolivia.

Gráfico 3. Dependencia parcial de la inflación mensual respecto de los regresores



Fuente: Elaboración propia, 2024

El gráfico 3 muestra que el IGAE tiene una relación débil con el IPC, sugiriendo que las variaciones en la actividad económica global no impactan significativamente la inflación mensual en Bolivia en el corto plazo. Esto indica que la actividad económica juega un rol menor en las fluctuaciones del IPC en el periodo analizado. En cambio, el IPC muestra una respuesta positiva ante aumentos en la tasa de interés efectiva, lo que sugiere que en Bolivia los incrementos en la tasa de interés están asociados con mayores presiones inflacionarias. Esto podría deberse al aumento de los costos de financiamiento, que se trasladan al consumidor final.

La relación entre la base monetaria y el IPC es negativa, lo que indica que aumentos en la base monetaria están asociados con una leve disminución en el IPC. Esto podría deberse a que el incremento de la base monetaria en Bolivia se dirige a sectores que no generan presión directa sobre los precios o al respaldo de las reservas, que estabilizan el efecto de la expansión monetaria en la inflación.

La oferta monetaria M2 tiene una relación negativa con el IPC, lo que sugiere que las políticas monetarias y cambiarias, como los controles de precios y la estabilidad cambiaria, moderan su impacto inflacionario. En cambio, el encaje legal muestra una relación positiva con el IPC, indicando que mayores encajes pueden aumentar la inflación al restringir el crédito y aumentar los costos de financiamiento.

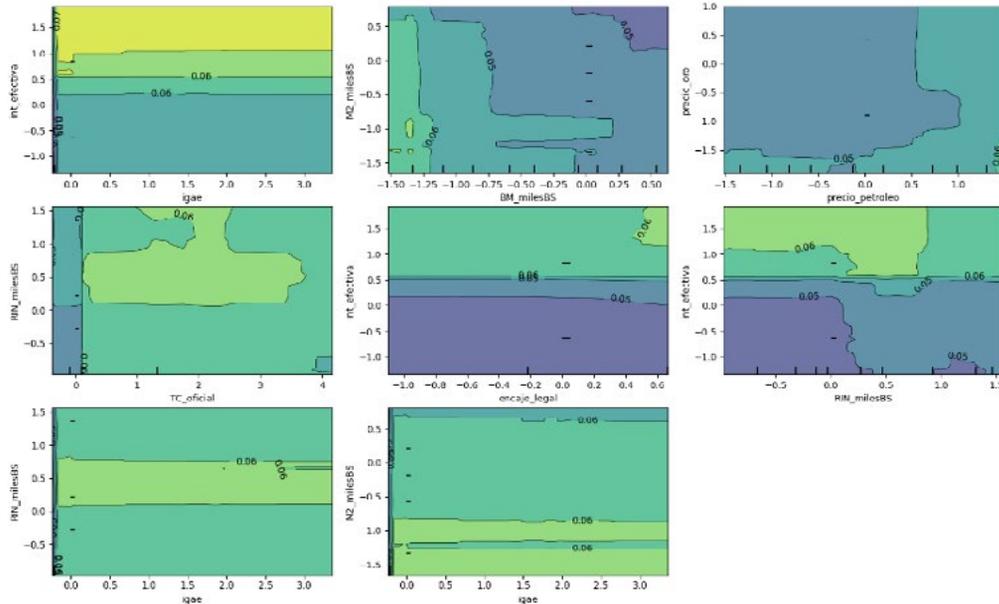
Los precios del petróleo, trigo y gas tienen una correlación positiva con el IPC, presionando la inflación por su impacto en los costos de producción y transporte. El precio del oro no afecta significativamente el IPC. El tipo de cambio oficial también está positivamente correlacionado con el IPC, ya que una depreciación de la moneda local encarece las importaciones. Las reservas internacionales muestran una correlación positiva moderada con el IPC, sugiriendo efectos indirectos de una política monetaria expansiva en la inflación.

En resumen, el análisis de dependencia parcial revela que la inflación en Bolivia está significativamente influenciada por factores externos, como el precio del trigo y del gas, y por una posible variación del tipo de cambio oficial. Estos factores reflejan la naturaleza importadora de la economía boliviana, donde los precios internacionales y la estabilidad cambiaria son determinantes clave en la dinámica inflacionaria. A su vez, las reservas internacionales juegan un papel en la política monetaria, aunque su impacto en la inflación es relativamente menor en comparación con otras variables.

Dependencia parcial de la inflación respecto a combinaciones de regresores

El gráfico 4, muestra el impacto conjunto de pares de variables independientes sobre la inflación mensual (IPC) en Bolivia, permitiendo identificar interacciones entre los regresores y su influencia combinada en la variación del IPC. Estos gráficos facilitan la interpretación de la forma en que dos factores pueden influir en la inflación al mismo tiempo, mientras los otros factores se mantienen constantes.

Gráfico 4. Dependencia parcial de la inflación mensual respecto de distintas combinaciones de regresores



Fuente: Elaboración propia, 2024

En el primer recuadro del gráfico 4 (de izquierda a derecha), que representa la interacción entre el IGAE y la tasa de interés efectiva, se observa una región donde el IPC muestra valores elevados cuando el IGAE es bajo y la tasa de interés es relativamente alta. Esto puede sugerir que, en escenarios de baja actividad económica, un aumento en la tasa de interés tiende a incrementar la presión sobre la inflación, posiblemente debido a mayores costos de financiamiento, lo cual contrasta con la expectativa clásica de que tasas más altas contienen la inflación. En el contexto boliviano, este comportamiento puede estar influido por la estructura de financiamiento en sectores específicos de la economía.

El análisis de la base monetaria y el M2 indica que la inflación aumenta cuando la base monetaria es alta y M2 se mantiene bajo, lo que sugiere que un exceso de liquidez en el sistema financiero puede impulsar los precios al alza, especialmente si no se acompaña de una expansión crediticia. En Bolivia, este exceso de liquidez puede reflejarse en mayores precios de consumo. El análisis de la interacción entre el precio del petróleo

y el encaje legal revela que la inflación tiende a aumentar con precios elevados del petróleo, independientemente del nivel del encaje legal. Esto resalta el impacto de los precios internacionales del petróleo en los costos de producción y transporte. Por otro lado, la relación entre el tipo de cambio oficial y las reservas internacionales netas muestra que una depreciación de la moneda local eleva el IPC debido al encarecimiento de las importaciones. Aunque las reservas internacionales contribuyen a estabilizar el tipo de cambio, su influencia directa sobre la inflación es limitada, especialmente en contextos de depreciación prolongada.

Por último, la interacción entre M2 e IGAE muestra que el IPC tiende a ser más estable cuando el IGAE está alto y M2 es bajo, lo que sugiere que una actividad económica fuerte junto con una oferta monetaria restringida ayuda a mantener la inflación controlada. Este escenario refleja una economía capaz de absorber aumentos de producción sin generar presiones inflacionarias significativas, lo que podría ser relevante en un contexto de control monetario. En general, los gráficos destacan cómo la interacción de variables económicas como el tipo de cambio, los precios de materias primas y la política monetaria influye en la inflación en Bolivia.

Estimación del Modelo Árbol de Decisión

Para estimar el modelo, se utilizó el Índice de Precios al Consumidor (IPC) como variable dependiente, siendo el objetivo de predicción. Las variables independientes incluyeron el resto de las variables del conjunto de datos, que aportan información relevante para modelar el IPC. Los datos se dividieron en tres subconjuntos: entrenamiento (165 observaciones), validación (20 observaciones) y testeo (22 observaciones). El conjunto de entrenamiento ajustó el modelo, el de validación se utilizó para optimizar hiperparámetros, y el de testeo se empleó para evaluar el rendimiento final del modelo.

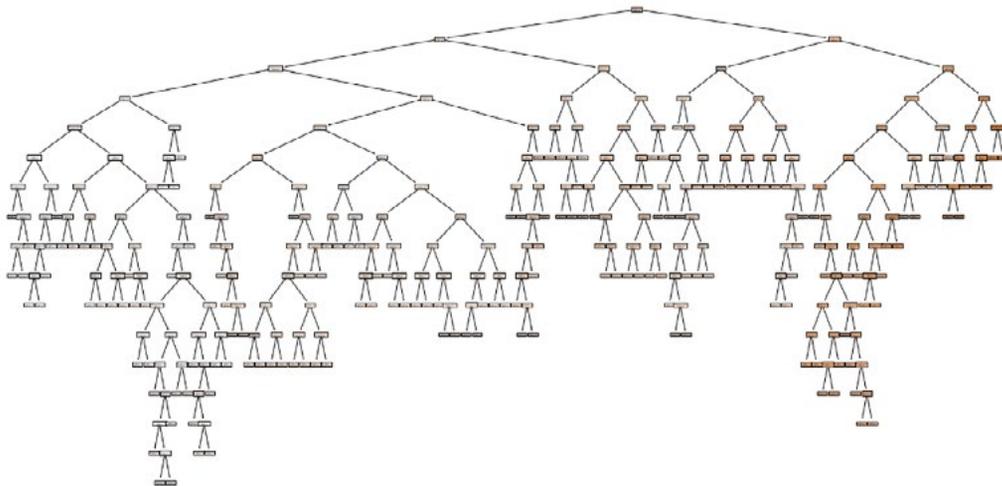
Tabla 2. Subconjuntos de entrenamiento, validación y testeo

Conjunto	Observaciones	Uso
Entrenamiento	165	Ajustar el modelo
Validación	20	Ajustar hiperparámetros en caso de optimización adicional
Testeo	22	Evaluación final del rendimiento del modelo
Total	207	

Fuente: Elaboración propia, 2024

No se estableció un límite predeterminado para la profundidad máxima del árbol, permitiendo que este crezca hasta que las hojas sean puras o contengan menos muestras de las necesarias para una división adicional. Se adoptó un valor mínimo de 1 muestra por hoja, permitiendo que cada hoja tuviera al menos una muestra. El criterio de división fue el error cuadrático medio, buscando minimizar el error en cada división del árbol. Además, el número mínimo de muestras para dividir un nodo fue 2, lo que significa que un nodo se divide si tiene al menos dos muestras.

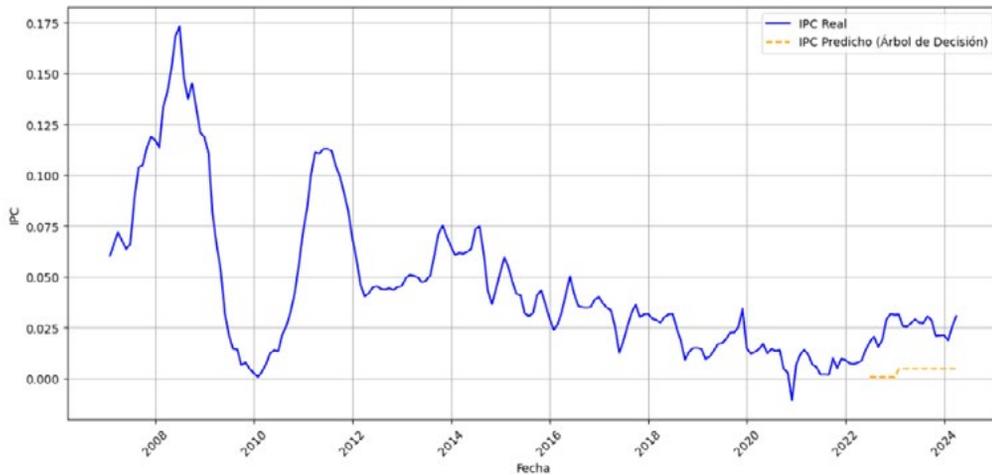
Gráfico 5. Árbol de decisión



Fuente: Elaboración propia, 2024.

El gráfico 5 muestra el árbol de decisión para predecir el IPC, donde los nodos dividen los datos en subconjuntos homogéneos y las hojas indican valores promedio del IPC. El modelo alcanzó un RMSE de 0.0223, aproximadamente la mitad de la variación promedio del IPC (0.0440), evidenciando su buena precisión.

El RMSE de 0.0223 es bajo en comparación con las fluctuaciones típicas del IPC, lo que sugiere que el modelo de árbol de decisión es efectivo para capturar las tendencias y variaciones del IPC durante el periodo analizado. En conclusión, el modelo muestra un rendimiento prometedor en la predicción de la variación mensual del IPC, con una precisión adecuada según el RMSE obtenido.

Gráfico 6. Comparación del IPC Real vs Predicho (Árbol de decisión)

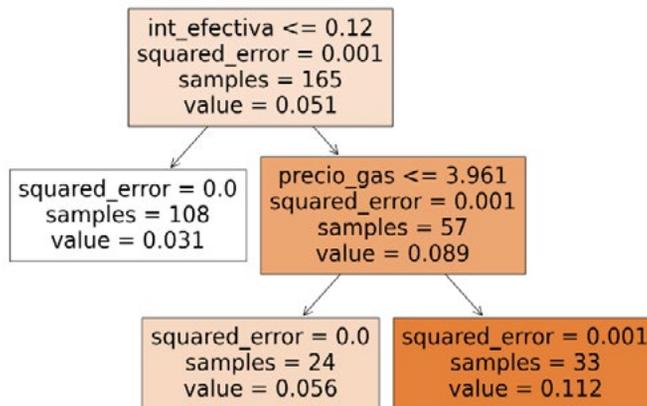
Fuente: Elaboración propia, 2004.

El gráfico 6 muestra que las predicciones del modelo de árbol de decisión coinciden estrechamente con los valores observados del IPC, con un RMSE de 0.0223, indicando alta precisión en la predicción del IPC.

Estimación del Modelo Árbol de Decisión Podado

El modelo de árbol de decisión podado se optimiza mediante la selección del parámetro de complejidad de costo (*ccp_alpha*), el cual regula la poda para mejorar la precisión y simplicidad del árbol. El *ccp_alpha*, o poda basada en complejidad de costo, penaliza el crecimiento del árbol eliminando ramas poco relevantes (Pedrogosa et al., 2011). Al incrementar el valor de *ccp_alpha*, la poda se vuelve más estricta, resultando en un árbol más pequeño y menos propenso al sobreajuste.

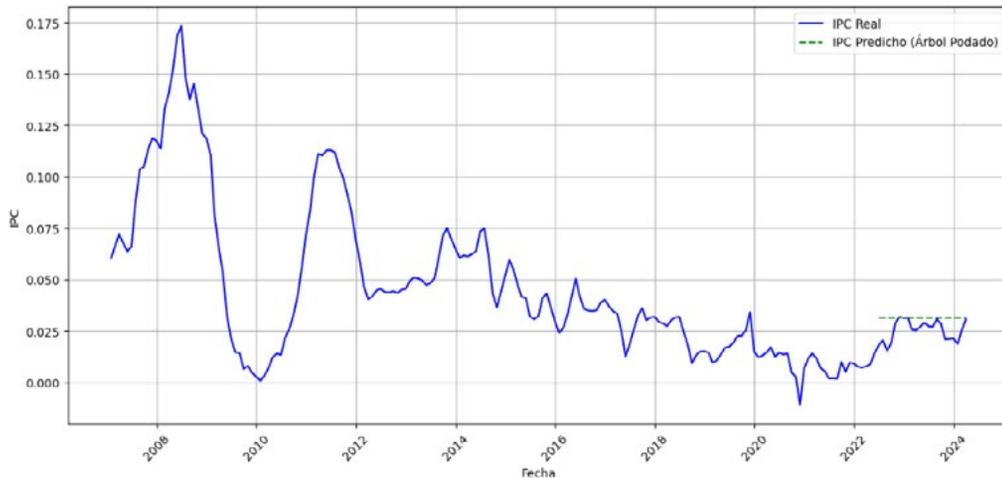
El valor óptimo de *ccp_alpha* de $9.96e-05$ minimizó el error y evitó el sobreajuste, logrando un RMSE de 0.0079 en el conjunto de testeo. Esto muestra una alta precisión en la predicción del IPC en Bolivia y un buen ajuste entre valores predichos y observados.

Gráfico 7. Árbol de Decisión Podado

Fuente: Elaboración propia, 2024

El gráfico 7 muestra el árbol de decisión podado, ajustado con el valor óptimo de *ccp_alpha* para reducir la complejidad sin sacrificar la capacidad predictiva. El modelo identifica dos variables clave en la predicción del IPC: la tasa de interés efectiva (*int_efectiva*) y el precio del gas (*precio_gas*). En el nodo raíz, se evalúa si la tasa de interés es menor o igual a 0.12. Cuando esto ocurre (rama izquierda), el IPC estimado es 0.031, con un error cuadrático de cero, lo que indica una predicción precisa y estable, sugiriendo que, en condiciones de baja tasa de interés, el IPC se mantiene moderado y estable.

En la rama derecha del árbol, cuando la tasa de interés efectiva supera 0.12, el modelo incorpora el precio del gas como factor clave. Si el precio del gas es inferior o igual a 3.961, el IPC estimado es 0.056, con un error cuadrático de cero. Sin embargo, si el precio del gas supera este umbral, el IPC estimado aumenta a 0.112, con un pequeño error cuadrático de 0.001, lo que sugiere cierta variabilidad influenciada por la volatilidad del mercado o el aumento de los precios del gas. El árbol de decisión podado resalta la relevancia de la tasa de interés efectiva y el precio del gas como variables clave para predecir el IPC, simplificando la estructura para facilitar la interpretación y la predicción económica con datos limitados.

Gráfico 8. Comparación del IPC Real vs Predicho (Árbol Podado)

Fuente: Elaboración propia, 2024

El gráfico 8 muestra que el modelo de Árbol de Decisión Podado predice el IPC con precisión (RMSE de 0.0079) durante 2007-2024. Aunque el IPC fluctuó inicialmente, se estabilizó desde 2012. El modelo se alinea con el promedio del IPC, evitando sobreajuste y reaccionando menos a pequeñas fluctuaciones, siendo útil para proyecciones a corto y mediano plazo.

En síntesis, el Árbol de Decisión Podado ofrece una estimación precisa del IPC, reflejada en su bajo RMSE, y captura de manera efectiva la tendencia general sin responder a variaciones menores. Esto hace que el modelo sea adecuado para aplicaciones en las que se desea un pronóstico estable y generalizado del IPC, manteniendo un bajo margen de error en relación con los valores históricos.

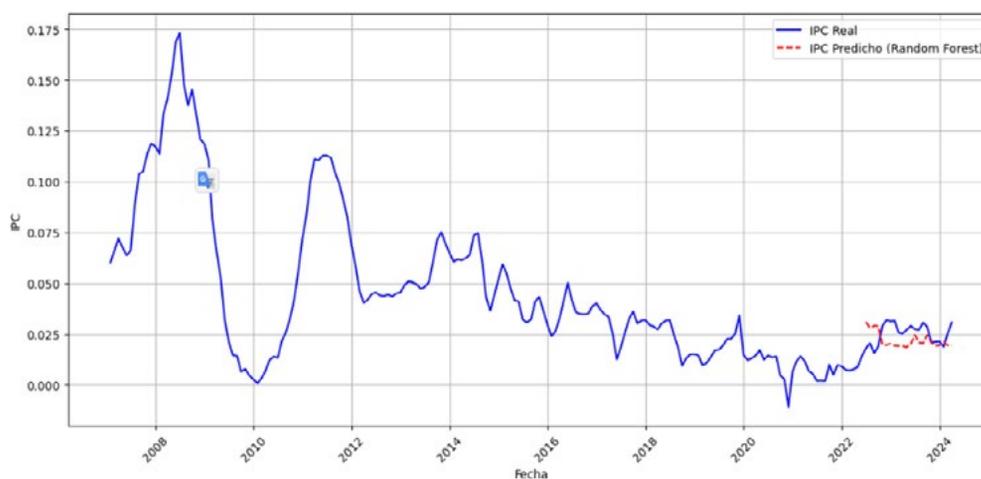
Estimación del Modelo Random Forest

Para predecir el IPC, se utilizó un modelo Random Forest con 100 árboles de decisión, equilibrando precisión y eficiencia computacional. El hiperparámetro *random_state* se fijó en 42 para asegurar la reproducibilidad, controlando la aleatoriedad en la selección de muestras y la partición de los datos. Esto permite obtener resultados consistentes en diferentes ejecuciones, facilitando la validación y comparación de pruebas. El modelo fue entrenado mediante $fit(x_{train}, y_{train})$, usando los datos históricos del IPC,

donde x_{train} contiene las variables predictoras y y_{train} la variable objetivo. Durante el entrenamiento, el modelo ajusta sus árboles para capturar patrones en los datos, aplicándolos luego en predicciones sobre nuevos datos.

El modelo Random Forest obtuvo un RMSE de 0.0085 en el conjunto de testeo, indicando que las predicciones están muy cercanas a los valores reales del IPC. Esta precisión, junto con su capacidad para manejar múltiples variables y detectar relaciones no lineales, demuestra su eficacia en la predicción del IPC en un entorno complejo. En el gráfico 9, la comparación entre valores reales y predichos muestra que el modelo sigue bien la tendencia general del IPC, aunque tiende a subestimar ligeras fluctuaciones extremas debido al efecto de suavizado al promediar múltiples árboles.

Gráfico 9. Comparación del IPC Real vs Predicho (Random Forest)



Fuente: Elaboración propia, 2024

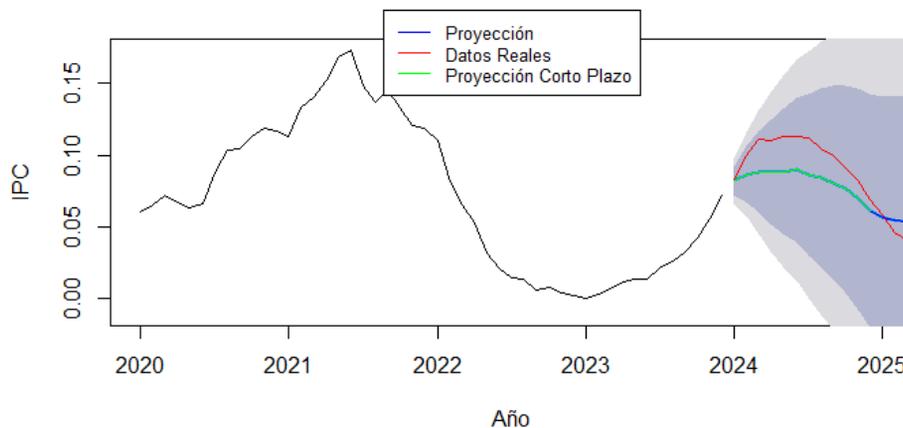
Estimación del Modelo ARIMA

Se estimó un modelo ARIMA(1,1,0)(1,0,0)[12] para predecir la variación mensual del IPC de enero de 2007 a marzo de 2024. Este modelo incluye un componente autoregresivo de primer orden ($p=1$), diferenciación de primer orden ($d=1$) para lograr estacionariedad, y un componente estacional autoregresivo de primer orden con periodicidad anual ($s=12$). Los coeficientes estimados son 0.6178 para el componente autoregresivo y -0.4991 para el estacional, con errores estándar de 0.1225 y 0.1215.

El modelo tiene un σ^2 de $6.266e-05$, un logaritmo de verosimilitud de 159.8, y criterios AIC, AICc, y BIC de -313.61, -313.05, y -308.05, respectivamente.

Las medidas de error del conjunto de entrenamiento mostraron una raíz del error cuadrático medio (RMSE) de 0.0077 y un error absoluto medio (MAE) de 0.0057. La función de autocorrelación del primer rezago (ACF1) fue de -0.0259. Finalmente, se realizó una predicción utilizando el modelo estimado, obteniéndose un RMSE de 0.0376 para el conjunto de datos de prueba. Este valor de RMSE indica que el modelo tiene un buen desempeño en la predicción de los datos de prueba, con errores relativamente pequeños.

Gráfico 10. Comparación del IPC real vs Predicho (modelo ARIMA)



Fuente: Elaboración propia, 2024

DISCUSIÓN

En este estudio, se compararon varios modelos de pronóstico del IPC en Bolivia, incluyendo enfoques de aprendizaje automático (Árbol de Decisión, Árbol de Decisión Podado y Random Forest) y un modelo tradicional ARIMA, evaluando su rendimiento mediante el RMSE en el conjunto de testeo.

Comparación entre el Árbol de Decisión, Árbol Podado y Random Forest

La Tabla 3 resume los resultados del RMSE obtenidos para los modelos de Árbol de Decisión, Árbol Podado y Random Forest. Observamos que el Árbol Podado presenta el RMSE más bajo en el conjunto de testeo (0.0079), lo cual indica que es el modelo con mejor precisión en la predicción del IPC. El modelo Random Forest también obtuvo un buen rendimiento, con un RMSE de 0.0085, apenas superior al del Árbol Podado, pero sustancialmente mejor que el Árbol de Decisión sin poda, cuyo RMSE alcanzó 0.0223. La poda en el árbol de decisión parece haber sido beneficiosa al reducir el sobreajuste, lo que resulta en una mayor capacidad predictiva del modelo en datos no observados.

Tabla 3. Comparación de los modelos de Árbol en términos de RMSE en el conjunto de testeo

Modelo	RMSE (conjunto de testeo)
Árbol de Decisión	0.0223
Árbol de Decisión Podado	0.0079
Random Forest	0.0085

Fuente: Elaboración propia, 2024

Estos resultados indican que el modelo Árbol Podado es el más adecuado entre las opciones de modelos de Árbol para el pronóstico del IPC. Si bien el Random Forest también ofrece una precisión elevada, el Árbol Podado logra un menor error de predicción, posiblemente debido a su simplicidad y la reducción de ruido que logra mediante la poda. A diferencia del Árbol de Decisión sin podar, ambos modelos –el Árbol Podado y el Random Forest– muestran mejor capacidad para generalizar en el conjunto de testeo.

Comparación entre el mejor Modelo de Aprendizaje Automático y el Modelo ARIMA

Tras identificar al Árbol Podado como el mejor modelo dentro de los enfoques basados en árboles, procedemos a comparar este modelo con el modelo tradicional ARIMA, cuyo RMSE fue de 0.0376 (ver Tabla 4). El modelo ARIMA muestra un desempeño inferior en términos de precisión en el conjunto de testeo, presentando un error significativamente mayor que el obtenido con los modelos de aprendizaje automático, especialmente el Árbol Podado.

Tabla 4. Comparación del mejor modelo de aprendizaje automático con el modelo ARIMA en términos de RMSE

Modelo	RMSE (conjunto de testeo)
Árbol de Decisión Podado	0.0079
ARIMA	0.0376

Fuente: Elaboración propia, 2024

La comparación destaca las ventajas del aprendizaje automático sobre enfoques tradicionales como ARIMA para predecir el IPC en Bolivia. Mientras que ARIMA es eficaz para series temporales, los modelos de aprendizaje automático, como el Árbol Podado, pueden capturar patrones más complejos y no lineales. El Árbol Podado, al aplicar poda basada en la complejidad del costo, mejoró la generalización del modelo y redujo el sobreajuste, lo cual es crucial en series económicas volátiles y variables.

En consecuencia, los resultados indican que el Árbol Podado es una alternativa eficiente y precisa para pronosticar la inflación en Bolivia, superando al modelo ARIMA en precisión (menor RMSE) y capturando mejor la dinámica del IPC. Aunque el Random Forest también demuestra robustez al utilizar múltiples árboles, el desempeño del Árbol Podado destaca por su simplicidad y efectividad. Se sugiere que investigaciones futuras exploren combinaciones híbridas de modelos de aprendizaje automático y métodos tradicionales para mejorar la precisión del pronóstico en economías emergentes.

CONCLUSIONES

Este estudio comparó la efectividad de modelos de pronóstico del Índice de Precios al Consumidor (IPC) de Bolivia, evaluando enfoques tradicionales como el modelo ARIMA y métodos de aprendizaje automático (Árbol de Decisión, Árbol Podado y Random Forest). Los resultados muestran que los modelos de aprendizaje automático, especialmente el Árbol de Decisión Podado, superan al modelo ARIMA en precisión, con un menor error cuadrático medio de raíz (RMSE) en el conjunto de testeo.

El Árbol Podado fue el modelo más preciso para pronosticar el IPC, con un RMSE de 0.0079, seguido por Random Forest con un RMSE de 0.0085. Ambos modelos superaron al Árbol de Decisión sin podar y al modelo ARIMA, cuyo RMSE fue mucho más alto (0.0376). Estos resultados indican que los modelos de aprendizaje automático son efectivos para pronosticar la inflación en economías emergentes como la boliviana, donde las dinámicas de precios son volátiles y no lineales.

La superioridad de los modelos de aprendizaje automático, especialmente el Árbol Podado, sobre el modelo ARIMA destaca la necesidad de considerar metodologías alternativas en el análisis económico. A diferencia de ARIMA, que tiene limitaciones para capturar relaciones no lineales, los modelos de árboles dividen los datos jerárquicamente según múltiples variables, ofreciendo una representación más precisa de las dinámicas subyacentes del IPC boliviano y adaptándose mejor a sus características específicas.

Este estudio muestra que los modelos de aprendizaje automático, como el Árbol Podado, son más precisos y adaptables que los enfoques tradicionales como el ARIMA para anticipar la inflación en Bolivia, representando un avance en el pronóstico económico.

RECOMENDACIONES

Si bien el modelo Árbol Podado demostró ser el más adecuado para el pronóstico del IPC en este estudio, se recomienda que futuras investigaciones exploren enfoques híbridos que combinen modelos de aprendizaje automático y técnicas tradicionales. Además, el uso de modelos de series temporales basados en redes neuronales, como las redes neuronales recurrentes (RNN) o las redes de memoria a largo plazo (LSTM), podría mejorar aún más la precisión del pronóstico al capturar dependencias temporales complejas. También sería relevante realizar estudios similares en otros países de América Latina para evaluar la aplicabilidad de estos modelos en diferentes contextos económicos y comparar los resultados obtenidos.

REFERENCIAS BIBLIOGRÁFICAS

1. Banco Central Europeo. (2023). *Quantile Regression Forests for Inflation Prediction in the Eurozone: Short-Term Adaptability in Forecasts*. Banco Central Europeo.
2. Bojanic, A. N. (2021). A Markov-Switching Model of Inflation in Bolivia. *Economies*.
3. Box, G. E., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
4. Bratu, M., & Nicula, A. S. (2024). Sentiment analysis as an innovation in inflation forecasting in Romania. *Marketing i menedžment inovacij*, 15(2), 13-25.
5. Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
6. Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC.
7. Carmona Rodríguez, C. S., & López Sánchez, A. (2021). *Revisión de implementación de modelo de Random Forest en los últimos años*. Universidad Pontificia Bolivariana.
8. Diaf, S., & Schütze, F. (2024). Economic forecasting with non-specific Google Trends sentiments Insights from US Data. *6th International Conference on Advanced Research Methods and Analytics* (págs. 10-17). Valencia: Universidad Politécnica de Valencia.
9. Espinosa, & Zúñiga. (2020). Aplicación de algoritmos Random Forest y XGBoost en la predicción de la inflación en México. *Ingeniería, Investigación y Tecnología*, 1-16.
10. Forte, A. (Septiembre de 2024). *Inflación de corto plazo con random forest (Documento de trabajo No. 24/10)*. BBVA Research. Obtenido de file:///D:/Courses%20taken/Pron%C3%B3sticos%20Macro%20Fin/Modulo%205.%20Taller%20de%20investigacion/Papers/A.%20Random-forest%20Argentina.pdf
11. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
12. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

13. Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning models. *Journal of Business & Economic Statistics*, 39(1), 98-119.
14. Medina Merino, R. F., & Ñique Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación para la predicción de la inflación en Perú. *Inter Fases*, 10, 165-180.
15. Meuller, S. (2022). Improving Inflation Forecast Accuracy with Random Forest Models. *Universidad de Lund, Departamento de Economía*.
16. Moyo, T., & Musengezi, J. (2022). Utilizing decision trees and random forests for inflation forecasting in African economies. *Journal of Economic Forecasting*, 58(3), 345-359.
17. Patton, A. J., & Simsek, Y. (2023). Generalized autoregressive score trees and forests in economic predictions. *International Journal of Forecasting*, 41(4), 221-238. .
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
19. Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81-106.
20. Silva, R. F., & Oliveira, M. A. (2019). Uso de Random Forest para la predicción de la inflación en Brasil. *Revista Brasileira de Agroinformática*, 15(2), 45-58.
21. Zhao, L., & Zhang, X. (2019). A Comparison of ARIMA and LSTM in Forecasting Time Series. *IEEE Xplore*.
22. Zhou, S., & Mentch, L. (2023). Trees, forests, and optimal pruning techniques in inflation forecasting. *Econometric Reviews*, 42(5), 634-657.