ISSN No. 2631-2743

KAIRÓS

Revista de ciencias económicas, jurídicas y administrativas

KAIRÓS, Vol. (9) No. 17, pp. 93 - 117, julio - diciembre 2026

MODELOS DE PREDICCIÓN ECONÓMICA UTILIZANDO APRENDIZAJE AUTOMÁTICO

ECONOMIC PREDICTION MODELS USING MACHINE LEARNING

Estéfany A. Orbe-Rosario
eaorbe2@espe.edu.ec
Universidad de las Fuerzas Armadas ESPE
(Quito - Ecuador)

Tatiana L. Palacios-Sinchi
tlpalacios@espe.edu.ec
Universidad de las Fuerzas Armadas ESPE
(Quito - Ecuador)

Pablo Gaibor Costta
pagaibor@espe.edu.ec
Universidad de las Fuerzas Armadas ESPE
(Quito - Ecuador)

DOI:
https://doi.org/10.37135/kai.03.17.05

Recibido: 08/12/2025
Aceptado: 15/06/2026

Unach
UNIVERSIDAD NACIONAL DE CHIMBORAZO
FACULTAD DE CIENCIAS POLÍTICAS Y ADMINISTRATIVAS

KAIRÓS
Revista de ciencias económicas, jurídicas y administrativas
ISSN No. 2631-2743

KAIRÓS, Vol. (9) No. 17, pp. 93 - 117, julio - diciembre 2026

MODELOS DE PREDICCIÓN ECONÓMICA UTILIZANDO APRENDIZAJE AUTOMÁTICO

ECONOMIC PREDICTION MODELS USING MACHINE LEARNING

DOI:
https://doi.org/10.37135/kai.03.17.05

Resumen

Este estudio analiza la capacidad predictiva de modelos econométricos y de aprendizaje automático para estimar el PIB real del Ecuador, con datos trimestrales de 2008-2022 e inflación y tasa de interés como variables explicativas. Se estimaron regresiones lineales múltiples mediante Mínimos Cuadrados Generalizados, modelos de rezagos distribuidos, Random Forest y redes neuronales LSTM. El desempeño predictivo se evaluó con métricas de error fuera de muestra, RMSE y MAE. Los resultados muestran que, en el contexto de alta volatilidad de 2022, el modelo de rezagos distribuidos optimizado obtuvo el menor error, mientras que en un escenario más estable el LSTM logró mayor precisión. Se concluye que la efectividad predictiva depende del modelo y del contexto económico, y que los enfoques econométricos tradicionales y las técnicas de aprendizaje automático pueden ser complementarios para la predicción macroeconómica en economías dolarizadas.

Palabras clave: pronósticos, series de tiempo, macroeconomía, econometría, redes neuronales

Abstract

This study analyzes the predictive capacity of econometric and machine learning models to estimate Ecuador's real GDP, using quarterly data for 2008-2022 and considering inflation and the interest rate as explanatory variables. Multiple linear regression models were estimated through Generalized Least Squares, along with distributed lag models, Random Forest, and Long Short-Term Memory (LSTM) neural networks. Predictive performance was assessed using out-of-sample error metrics, namely RMSE and MAE. The results show that, in the highly volatile context of 2022, the optimized distributed lag model achieved the lowest prediction error, whereas in a more stable macroeconomic scenario the LSTM model reached higher relative accuracy. The study concludes that predictive effectiveness depends on both the model specification and the economic context, and that traditional econometric approaches and machine learning techniques may be complementary for macroeconomic forecasting in dollarized economies.

Keywords: forecasting, time series, macroeconomics, econometrics, neural networks

Introducción

La predicción de variables macroeconómicas constituye una herramienta fundamental para la formulación de políticas públicas, la planificación económica y la toma de decisiones en el sector privado. Indicadores como el Producto Interno Bruto (PIB), la inflación y las tasas de interés permiten anticipar escenarios económicos, evaluar riesgos y diseñar estrategias que mitiguen los efectos de ciclos adversos o crisis económicas. En economías dolarizadas como la ecuatoriana, donde la política monetaria es limitada, la capacidad de anticipar el comportamiento de estas variables adquiere especial relevancia.

Tradicionalmente, la predicción económica ha sido abordada mediante modelos econométricos clásicos, basados en relaciones causales y supuestos lineales. No obstante, la creciente complejidad de los sistemas económicos ha evidenciado limitaciones en estos enfoques para capturar dinámicas no lineales y patrones cambiantes en los datos. En este contexto, el desarrollo tecnológico ha ampliado las posibilidades de análisis. El auge de las redes y aplicaciones informáticas ha permitido optimizar procesos y proponer soluciones a problemas complejos mediante herramientas computacionales (Dariel & Rainiero, 2019).

Estas herramientas pueden aprender a partir de los datos, identificar patrones de comportamiento y estimar relaciones entre variables, reduciendo la dependencia de análisis puramente intuitivos que podrían carecer de respaldo empírico (Espino, Martínez, & Daradoumis, 2017). En el caso ecuatoriano, la volatilidad observada en distintos periodos económicos y las limitaciones estructurales de una economía dolarizada refuerzan la necesidad de contar con herramientas de predicción confiables. Sin embargo, la evidencia empírica comparativa entre modelos econométricos tradicionales y modelos de aprendizaje automático en el contexto nacional sigue siendo limitada, lo que evidencia una brecha en la literatura aplicada.

A diferencia de estudios previos centrados exclusivamente en la comparación entre modelos econométricos y técnicas de aprendizaje automático, la presente investigación incorpora un análisis diferenciado según el contexto macroeconómico, evaluando el desempeño predictivo de los modelos tanto en periodos de estabilidad como en escenarios de alta incertidumbre económica. Aplicado al caso ecuatoriano, caracterizado por una economía dolarizada y vulnerable a choques externos, este enfoque permite analizar cómo las condiciones macroeconómicas afectan la precisión relativa de los distintos modelos predictivos. De esta manera, el estudio aporta evidencia empírica relevante para la predicción macroeconómica en economías emergentes y dolarizadas.

En consecuencia, resulta pertinente desarrollar herramientas que permitan predecir el comportamiento de variables macroeconómicas como el PIB, la inflación y las tasas de interés, indicadores clave para la toma de decisiones tanto en el ámbito público como privado. Con ello, se busca proveer a los hacedores de política y planificadores económicos una herramienta empírica robusta que reduzca el sesgo de proyección en entornos de eventos extraordinarios.

El presente trabajo se enfoca en el desarrollo y aplicación de modelos econométricos clásicos y modelos de aprendizaje automático supervisado con el

fin de evaluar su desempeño en la predicción de variables macroeconómicas. En particular, la investigación busca responder a la siguiente pregunta central: ¿Qué tipos de modelos son más precisos para predecir el comportamiento de variables macroeconómicas?

El análisis se realiza considerando dos escenarios diferenciados: un periodo de relativa estabilidad macroeconómica y un periodo que incorpora episodios de alta incertidumbre, con el fin de evaluar la robustez predictiva de cada enfoque. Asimismo, se examinan ventajas y limitaciones de los modelos en el contexto ecuatoriano, de manera que los resultados puedan servir como referencia para futuros estudios en economías dolarizadas con características similares.

Metodología

Fuentes y técnicas de recolección de datos

La investigación se fundamenta en datos secundarios de carácter cuantitativo, provenientes de fuentes oficiales nacionales e internacionales. En particular, se utilizaron datos de las siguientes fuentes:

- El Banco Central del Ecuador (BCE), principal fuente de datos macroeconómicos nacionales.
- El Instituto Nacional de Estadística y Censos (INEC), que proporciona información sociodemográfica y económica complementaria,
- Organismos internacionales como el Fondo Monetario Internacional (FMI) y el Banco Mundial, que ofrecen datos macroeconómicos estandarizados a nivel global.

La base de datos empleada fue consolidada mediante procesos de extracción, depuración y organización de información proveniente de estas fuentes. Esta base constituye el insumo principal para el entrenamiento y evaluación de los modelos predictivos. El periodo de análisis comprende los años 2008 – 2022, lo que permite abarcar distintas fases del ciclo económico ecuatoriano, incluyendo episodios de estabilidad y de alta incertidumbre macroeconómica. La frecuencia de los datos es trimestral, seleccionada en función de su disponibilidad y de la necesidad de contar con suficiente granularidad para el modelado de series temporales.

Desde el punto de vista del propósito, se trata de una investigación explicativa, correlacional y predictiva. Es explicativa porque pretende identificar causas y efectos dentro de fenómenos económicos complejos, como los factores que inciden en la variación del PIB. Es correlacional porque busca medir la relación entre variables económicas como las tasas de interés y la inflación, y cómo estas influyen entre sí. Por último, es predictiva puesto que uno de los principales objetivos es desarrollar modelos que puedan anticipar el comportamiento futuro de variables económicas, a partir de patrones históricos.

En cuanto al diseño metodológico, se trata de un estudio de tipo cuantitativo, ya que se aplicaron diferentes técnicas de modelado y simulación para analizar

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026
(Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

el efecto de distintas combinaciones de variables independientes sobre el comportamiento de la variable dependiente. Este estudio se realizará mediante la construcción y prueba de modelos predictivos utilizando métodos econométricos, estadísticos y de aprendizaje automático. A través de esta aproximación se busca simular escenarios, evaluar hipótesis y seleccionar los modelos con mejor capacidad explicativa y predictiva.

El enfoque se considera longitudinal, ya que se analiza información a lo largo del tiempo, permitiendo observar tendencias y cambios estructurales. También se considera observacional en el sentido de que los datos utilizados no son generados artificialmente, sino que provienen de fuentes oficiales y reflejan fenómenos reales. El marco metodológico que sustenta el presente trabajo se basa en una combinación entre los principios de la teoría económica clásica y moderna, junto con los métodos contemporáneos de la ciencia de datos, como la estadística avanzada, el modelado matemático y el aprendizaje automático.

Variables

Variable dependiente

La principal variable que se busca predecir es el PIB real, por tratarse de uno de los indicadores más representativos del desempeño macroeconómico. De manera complementaria, se estiman modelos para otras variables macroeconómicas relevantes.

Variables independientes

Las variables incorporadas en los modelos corresponden a indicadores macroeconómicos seleccionados por su relevancia teórica y empírica para explicar el comportamiento del PIB. En particular, se consideraron la inflación y la tasa de interés activa referencial, así como los rezagos del PIB y de las variables explicativas en aquellas especificaciones donde la estructura del modelo lo requirió. Estas variables se seleccionan por su capacidad para capturar dinámicas temporales que contribuyan a mejorar la precisión predictiva de los modelos, más que por su interpretación causal directa (Hyndman & Athanasopoulos, 2021).

La selección de variables respondió tanto a criterios teóricos como a restricciones de disponibilidad y consistencia estadística de las series trimestrales. La inflación y la tasa de interés fueron priorizadas debido a su relevancia macroeconómica y a su capacidad para capturar dinámicas monetarias y financieras asociadas al comportamiento del PIB en economías dolarizadas como la ecuatoriana. Si bien existen otros factores relevantes para la economía ecuatoriana, como el gasto público, el comercio exterior, la inversión, el consumo interno o los precios del petróleo, su incorporación no fue considerada dentro de esta investigación debido al enfoque parsimonioso del modelo y con el propósito de evitar incrementos innecesarios en la complejidad, problemas de multicolinealidad y posibles riesgos de sobreajuste.

Adicionalmente, la incorporación de rezagos permitió capturar efectos

dinámicos e intertemporales relevantes para la predicción económica.

Preparación y procesamiento de datos

El tratamiento de los datos constituye una etapa fundamental del estudio, ya que la calidad del conjunto de datos afecta directamente la precisión de los modelos predictivos. La base principal proviene del Banco Central de Ecuador, y abarca el período comprendido entre los años 2008 y 2022, con frecuencia trimestral. Esta fuente se complementa con una base de datos interna consolidada por el equipo de investigación.

Con el propósito de evaluar la estabilidad y robustez predictiva de los modelos bajo distintos contextos macroeconómicos, el estudio consideró dos escenarios de análisis diferenciados. El primero corresponde a un periodo de relativa estabilidad económica, utilizando información de entrenamiento entre 2008 y 2018 y datos de prueba para el año 2019. El segundo escenario incorpora un contexto de disrupciones económicas asociado a la pandemia y al paro nacional, utilizando datos de entrenamiento entre 2008 y 2021 y evaluación fuera de muestra para el año 2022. Esta estrategia permitió comparar el desempeño de los distintos modelos tanto en entornos estables como en escenarios de alta volatilidad.

La primera etapa del procesamiento consistió en la depuración de la base de datos, incluyendo la revisión de valores faltantes y atípicos. En presencia de datos faltantes se emplearon técnicas de interpolación temporal para preservar la continuidad de las series (Hyndman & Athanasopoulos, 2021).

Posteriormente, se realizó un proceso de ingeniería de características (feature engineering), generando rezagos temporales y transformaciones logarítmicas de las variables. Estas transformaciones permiten capturar dinámicas intertemporales y estabilizar la varianza de las series, lo cual resulta útil en modelos de series de tiempo (Brownlee, 2020).

Para los modelos de aprendizaje automático (Random Forest y LSTM) se aplicó normalización Min-Max, con el fin de evitar que las diferencias de escala entre variables afecten el proceso de entrenamiento. En contraste, en los modelos econométricos se emplearon transformaciones logarítmicas y especificaciones dinámicas sin requerir escalamiento previo.

Se efectuó un análisis exploratorio mediante gráficos de series de tiempo y análisis de correlación, con el propósito de identificar tendencias, posibles cambios estructurales y relaciones preliminares entre variables.

Finalmente, para validar la consistencia estadística de los modelos econométricos se aplicaron pruebas complementarias. El test de Breusch-Pagan se utilizó para detectar posibles problemas de heterocedasticidad, mientras que el estadístico Durbin-Watson permitió evaluar la presencia de autocorrelación en los residuos. Asimismo, se empleó el factor de inflación de la varianza (VIF) para identificar problemas de multicolinealidad entre variables explicativas. Finalmente, el desempeño predictivo de los modelos se comparó mediante las métricas RMSE (Root Mean Squared Error) y MAE (Mean Absolute Error), las cuales permiten cuantificar la magnitud promedio de los errores de predicción fuera de muestra.

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026
(Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

Modelos econométricos utilizados

A fin de analizar las variables macroeconómicas y evaluar su capacidad predictiva, el presente trabajo incorpora modelos de predicción econométricos tradicionales, así como técnicas modernas de aprendizaje automático. Particularmente se emplea la regresión lineal múltiple MCO, un modelo de regresión dinámica denominado Autorregresivo de Rezagos distribuidos ARDL y en cuanto a los modelos de aprendizaje automático, se utiliza bosque aleatorio, conocido en inglés como *Random Forest Regressor*, y modelos de redes de memoria de corto y largo plazo, conocidas como LSTM (*Long Short-Term Memory*).

La combinación de los modelos mencionados permitirá evaluar la precisión de las variables macroeconómicas, PIB, inflación y tasas de interés, así como analizar diferencias en cuanto a la interpretabilidad, complejidad, y su adaptabilidad de los distintos modelos a los datos temporales.

Regresión lineal múltiple

"La regresión lineal múltiple trata de ajustar modelos lineales entre una variable dependiente y más de una variable independiente. En este tipo de modelos es importante testar la heterocedasticidad, la multicolinealidad y la especificación", dicho modelo es el más simple (Montero, 2016). La ecuación general del modelo se expresa como:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (1)$$

En la ecuación Y_i representa la variable dependiente del estudio, mientras que X_{2i}, X_{3i}, … X_{ki} las variables explicativas. En este contexto, las variables a estimar son el PIB, inflación y tasas de interés respectivamente, por otro lado es el término constante o conocido también como intercepto, y coeficientes regresión parcial que indican el cambio esperado en la variable dependiente ante un cambio unitario en cada uno de los coeficiente de regresión, manteniendo constante la influencia del otro, por último como el término de perturbación estocástica, que para efectos del estudio se sustituye por para denotar series de tiempo obteniendo como ecuación final la siguiente (Gujarati & Porter, 2010).

$$Y_i = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad (2)$$

Modelo de rezagos distribuidos

El modelo de regresión dinámica con rezagos considera datos actuales y pasados de las variables explicativas, y en algunos casos, los valores rezagados de la variable dependiente. Este modelo es comúnmente usado en econometría para modelar series temporales, que permitan capturar los efectos retardados de las variables explicativas y también la inercia de la variable dependiente en el tiempo (Gujarati & Porter, 2010). La ecuación general del modelo de rezagos distribuidos es:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{(t-1)} + \cdots + \beta_k X_{(t-k)} + \mu_t \quad (3)$$

Donde Y_t representa la variable dependiente en el tiempo, X_t, X_{t-1} … X_{t-k} los valores actuales y rezagados de la variable explicativa, por otro lado β_0, β_1, … β_k son los coeficientes que miden el impacto de los valores actuales y rezagados de X; α es un intercepto y μ_t el término de error del modelo (Gujarati & Porter, 2010). Cuando el modelo incluye rezagos de la variable dependiente se denomina modelo autorregresivo, y adopta la siguiente estructura (Gujarati & Porter, 2010).

$$Y_t = \alpha + \beta X_t + \gamma Y_{(t-1)} + \mu_t \quad (4)$$

En este caso captura la dependencia de con respecto a su valor pasado Y_{t-1}, este , este modelo combina elementos de una regresión lineal estándar con características dinámicas, lo cual lo hace adecuado para modelar efectos económicos con retrasos temporales (Gujarati & Porter, 2010). En el contexto del presente estudio, el modelo propuesto incluye tanto rezagos de las variables explicativas, así como de la variable dependiente, permitiendo analizar los efectos a corto y largo plazo.

Si bien la especificación inicial de los modelos se basa en el enfoque de Mínimos Cuadrados Ordinarios (MCO), durante la fase de estimación se identificaron posibles problemas de heterocedasticidad, autocorrelación y endogeneidad, frecuentes en series macroeconómicas. Por esta razón, tras la evaluación de los supuestos del modelo clásico de regresión lineal, las estimaciones finales se realizaron mediante un enfoque de Mínimos Cuadrados Generalizados (GLS), asumiendo una estructura de autocorrelación de primer orden AR (1) en los errores.

Este procedimiento permite obtener estimadores más eficientes en presencia de autocorrelación serial, situación común en series de tiempo macroeconómicas como el PIB, la inflación y las tasas de interés.

Modelos de machine learning

Random forest

Es un método de aprendizaje automático basado en la combinación de múltiples árboles de decisión mediante técnicas de bagging y selección aleatoria de variables (Breiman, 2001). El bagging consiste en entrenar cada árbol con muestras Bootstrap del conjunto de datos, es decir, selecciones aleatorias con reemplazo. Esto implica que algunas observaciones pueden repetirse mientras que otras quedan fuera de la muestra de entrenamiento de cada árbol, conocidas como observaciones out-of-bag (OOB), las cuales pueden emplearse para estimar el error del modelo.

Adicionalmente, en cada nodo del árbol se selecciona de forma aleatoria un subconjunto de variables explicativas para determinar la mejor partición. Este proceso reduce la correlación entre árboles y mejora la capacidad de generalización

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026
(Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

del modelo. En términos generales, la predicción del modelo se expresa como:

$$\hat{Y} = 1/B \sum_{(b=1)}^{B} T_{\beta}(X) \quad (5)$$

Donde: $T_b(X)$ representa la predicción del árbol b-ésimo construido a partir de una muestra aleatoria del conjunto de datos y un subconjunto de variables explicativas (Breiman, 2001).

Para la implementación del modelo en esta investigación, se utilizó una partición temporal de los datos: el periodo 2008 – 2018 se destinó al entrenamiento y el año 2019 a la evaluación fuera de muestra. El modelo se estimó con 500 árboles y selección aleatoria de predictores en cada nodo. El desempeño predictivo se evaluó mediante RMSE y MAE, seleccionando el modelo con menor error fuera de muestra. Los hiperparámetros utilizados, particularmente el número de árboles (ntree=500) y la cantidad de variables seleccionadas aleatoriamente en cada partición (mtry=2), fueron definidos con base en pruebas experimentales y recomendaciones frecuentes en la literatura especializada, priorizando configuraciones que redujeran el error fuera de muestra y minimizaran riesgos de sobreajuste.

Long short-term memory (LSTM)

Las redes LSTM, o redes de memoria de corto y largo plazo, constituyen una variación avanzada de las redes neuronales recurrentes (RNN), diseñadas para modelar dependencias temporales de largo plazo en datos secuenciales. Estas redes permiten superar el problema del desvanecimiento del gradiente, frecuente en RNN tradicionales, lo que favorece un aprendizaje más estable de patrones complejos en series temporales (DataScientest, 2024).

El problema del desvanecimiento del gradiente surge durante la retropropagación a través del tiempo, cuando los gradientes tienden a disminuir exponencialmente a medida que se transmiten hacia periodos anteriores, generando variaciones mínimas en los pesos de las capas iniciales y dificultando el aprendizaje de relaciones de largo plazo (DataScientest, 2024).

Para mitigar este fenómeno, las LSTM incorporan celdas de memoria controladas por compuertas que regulan qué información se conserva, se descarta o se transmite en el tiempo. Este mecanismo permite preservar información relevante durante múltiples periodos y mejorar la capacidad del modelo para capturar dinámicas intertemporales en series económicas (DataScientest, 2024).

En esta investigación, el modelo LSTM se empleó para capturar relaciones no lineales y patrones dinámicos entre el PIB, la inflación y la tasa de interés activa referencial. Previamente al entrenamiento, las variables fueron normalizadas mediante escalamiento Min-Max con el fin de mejorar la estabilidad del proceso de aprendizaje. Se construyeron secuencias temporales de cuatro rezagos (n_steps = 4), de modo que cada observación de entrada contiene información de los cuatro trimestres previos del PIB, la inflación y la tasa de interés.

El modelo se configuró con una capa LSTM de 50 unidades ocultas, función de pérdida de error cuadrático medio (MSE), optimizador Adam con tasa de aprendizaje de 0.01 y 300 épocas de entrenamiento.

La muestra se dividió en conjunto de entrenamiento (periodo 2008 - 2021) y conjunto de prueba (2022). Para generar las secuencias del conjunto de prueba se utilizaron adicionalmente los últimos cuatro trimestres del periodo de entrenamiento, garantizando la continuidad temporal requerida por el modelo. Posteriormente, las predicciones fueron desnormalizadas para su interpretación en niveles originales. El desempeño del modelo se evaluó mediante el RMSE fuera de muestra.

La selección de las variables explicativas responde a su relevancia teórica y empírica en la determinación del PIB. En particular, la inflación y la tasa de interés activa referencial fueron consideradas debido a su incidencia directa sobre el consumo, la inversión y las condiciones de financiamiento en economías dolarizadas, donde la política monetaria presenta restricciones estructurales.

En cuanto a los modelos utilizados, la inclusión de regresiones lineales múltiples permite capturar relaciones contemporáneas entre variables macroeconómicas, mientras que los modelos de rezagos distribuidos incorporan efectos dinámicos y retardados en la evolución del PIB. Por su parte, la incorporación de algoritmos de aprendizaje automático, como Random Forest y redes neuronales LSTM, responde a su capacidad para modelar relaciones no lineales y patrones complejos en series temporales, lo que permite contrastar su desempeño frente a enfoques econométricos tradicionales.

Los hiperparámetros de los modelos de aprendizaje automático fueron definidos mediante criterios experimentales y revisión de literatura especializada, priorizando configuraciones parsimoniosas acordes con el tamaño de la muestra disponible. En el modelo Random Forest se utilizaron 500 árboles (ntree=500) y dos variables por partición (mtry=2), mientras que el modelo LSTM fue configurado con 50 unidades ocultas, secuencias de cuatro rezagos y un entrenamiento de 300 iteraciones completas sobre el conjunto de entrenamiento.

Estas configuraciones permitieron controlar la complejidad de los modelos y reducir riesgos de sobreajuste en presencia de una muestra relativamente limitada. Asimismo, se aplicó una validación temporal mediante separación entrenamiento-prueba, evitando utilizar información futura durante el entrenamiento y preservando la secuencia cronológica de las series. Finalmente, la división del análisis en distintos contextos económicos permite evaluar la robustez de los modelos bajo diferentes condiciones macroeconómicas, fortaleciendo la validez de los resultados obtenidos.

Tabla de configuración metodológica

Con el propósito de resumir la configuración aplicada en la estimación de los modelos predictivos, la Tabla 1 presenta los periodos de entrenamiento y prueba, variables utilizadas, métricas de evaluación y los principales parámetros considerados en cada especificación.

Tabla 1. Configuración metodológica de los modelos predictivos

Modelo	Periodo entrenamiento / prueba	Parámetros principales
Regresión lineal múltiple (Modelos 1, 2 y 3)	2008-2021 / 2022	Estimación por MCO y ajuste mediante GLS con estructura AR(1).

Modelos de rezagos distribuidos (Modelos 4 y 5)	2008-2021 / 2022	Inclusión de rezagos trimestrales; Modelo 5 como especificación optimizada.
Random Forest (Modelo 6)	2008-2021 / 2022	500 árboles (ntree=500) y dos variables por partición (mtry=2).
LSTM (Modelo 7)	2008-2021 / 2022	50 unidades ocultas, cuatro rezagos, optimizador Adam, tasa de aprendizaje 0.01 y 300 épocas.

Fuente: elaboración propia.

Resultados

El análisis y proyección del PIB constituyen una herramienta fundamental en el estudio macroeconómico, ya que permiten evaluar la evolución de los ciclos económicos agregados y comprender los determinantes del crecimiento económico. Anticipar su comportamiento facilita la evaluación de la capacidad productiva de la economía y la identificación de posibles escenarios de crisis o expansión.

El presente estudio emplea un enfoque cuantitativo para modelar y predecir el comportamiento del PIB del Ecuador para el año 2022 utilizando técnicas econométricas tradicionales y métodos de aprendizaje automático. La base de datos empleada cubre el periodo comprendido entre el primer trimestre del año 2008 al cuarto trimestre del año 2021, para tal efecto, el PIB será predicho en función de las variables explicativas, inflación y tasas de interés.

La metodología contempla la especificación de modelos de regresión lineal múltiple en distintas formas funcionales (lineal, log-lineal y con variables dummy para eventos atípicos, como la pandemia de 2020). Si bien la formulación inicial se plantea bajo el enfoque de Mínimos Cuadrados Ordinarios (MCO), las estimaciones finales se realizan mediante Mínimos Cuadrados Generalizados (GLS), incorporando una estructura autorregresiva de orden uno AR (1) en los errores, con el fin de corregir problemas de autocorrelación detectados en las pruebas diagnósticas. Asimismo, se incorporan modelos de rezagos distribuidos para captar efectos retardados.

Regresión lineal múltiple

Luego de un análisis de correlación, se construyeron 3 modelos de regresión lineal múltiple tradicional, bajo la metodología de mínimos cuadrados ordinarios, en los cuales se mantiene la misma estructura funcional: la variable dependiente es el PIB real (PIBc), y las variables explicativas, la inflación trimestral (inflación) y la tasa de interés (tiarv), resultando el siguiente modelo:

$$pib_{(c)} = \beta_0 + \beta_1 * \text{inflación} + \beta_2 * \text{tiarv} + \mu \quad (6)$$

Estos modelos se entrenaron con los datos de los años comprendidos entre 2008 y 2021, posteriormente se creó un modelo de prueba (test) con base en el año 2022, es decir las 4 observaciones que se excluyeron de la base de entrenamiento, con la finalidad de comparar si los valores predichos se asemejan a los valores reales.

Los modelos se ajustaron mediante mínimos cuadrados ordinarios, se evaluaron los supuestos del modelo clásico de regresión lineal haciendo énfasis en los posibles problemas de heteroscedasticidad y autocorrelación de los errores. Para todos los modelos, en el primer caso no se rechazó la hipótesis nula del test de Breusch-Pagan, concluyendo que los errores son homocedásticos. También se detectó autocorrelación mediante el test de Durbin Watson. En respuesta a esto, se generó un modelo con una estructura de mínimos cuadrados generalizados, asumiendo que los errores tienen una estructura de autocorrelación de grado 1 AR (1). Posteriormente se evaluó los coeficientes resultantes de cada modelo que se pueden visualizar en la Tabla 2.

Tabla 2. Coeficientes de los modelos de regresión lineal múltiple de los 3 modelos aplicados

Variables	Modelo 1	Modelo 2	Modelo 3
Intercepto	34666.47 (431444.2)	11.9292 (1856.55)	34681.54 (12563)
Inflación	- 19980.47 (31230.1)z	0.007818 ** (0.0039)	19534.24** (31230.6)
Tasa de interés	- 1237.15** (495.7)	0. 826568** (0. 2772)	-1238.24** (495.6)
Pandemia	-	-	-480.13 (478.0)
Observaciones	56	44	56
AIC	844.11	- 147.93	830.926

Nota. Nivel de significancia al 1% (***p<0.01), 5% (**p<0.05), 10% (*p<0.1). Los errores estándar robustos se muestran en paréntesis. La variable dependiente corresponde al PIB.
Fuente: elaboración propia.

Luego de haber estimado los modelos econométricos, se procedió a generar las predicciones del PIB correspondiente al año 2022. Para ello se utilizaron las funciones de predicción aplicadas sobre el conjunto de prueba, obteniendo así los valores esperados del PIB, bajo cada enfoque, que se pueden observar en la Tabla 3.

Tabla 3. Comparación entre PIB real y PIB predicho de los modelos aplicados.

Trimestre	PIB Real	PIB Predicho Modelo 1.	PIB Predicho Modelo 2.	PIB Predicho Modelo 3.
2022I	28198	22910	22451	22913
2022II	18558	23207	22942	23209
2022III	28685	23309	23035	23313
2022IV	29607	23228	22805	23233

Fuente: elaboración propia.

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026 (Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

Como se puede observar en la Tabla 3 las predicciones generadas por los tres modelos para los trimestres del año 2022 muestran diferencias importantes respecto a los valores reales del PIB. En general, los modelos tienden a subestimar el PIB en la mayoría de los trimestres, con excepción del segundo trimestre, donde se produce una sobreestimación. En el caso del modelo 1, las predicciones se mantienen relativamente estables alrededor de los 23 mil puntos, sin lograr capturar la variabilidad observada en el PIB real. Esto evidencia limitaciones del modelo para adaptarse a fluctuaciones pronunciadas en el periodo de evaluación.

En el modelo 2, estimado en forma logarítmica, presenta desviaciones aún mayores en varios trimestres. Si bien la transformación logarítmica puede mejorar la estabilidad en ciertos contextos, en este caso no se traduce en una mejora de la precisión predictiva, mostrando errores considerables frente a los valores reales. Por su parte, el modelo 3, que incorpora una variable dummy para capturar efectos estructurales, exhibe predicciones muy similares al modelo 1, pero con errores ligeramente menores en la mayoría de los trimestres. Esto sugiere que la inclusión del componente estructural aporta cierta mejora marginal en la capacidad predictiva.

Tabla 4. Métricas de precisión

	Modelo 1.	**Modelo 2.**	**Modelo 3.**
RMSE	5458.07	5710.15	5455.37
MAE	5422.84	5645.43	5420.39

Fuente: elaboración propia.

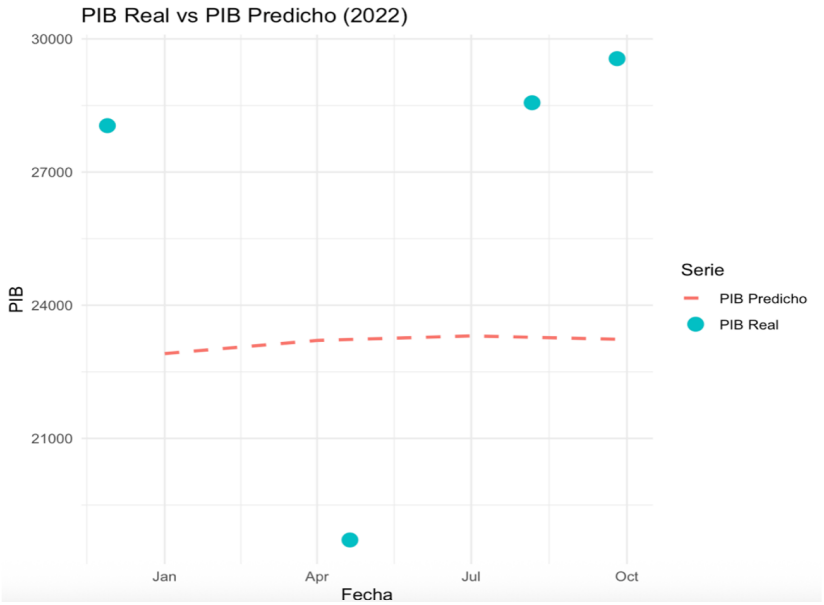
Como se puede observar en la Tabla 3 las predicciones generadas por los tres modelos para los trimestres del año 2022 muestran diferencias importantes respecto a los valores reales del PIB. En general, los modelos tienden a subestimar el PIB en la mayoría de los trimestres, con excepción del segundo trimestre, donde se produce una sobreestimación. En el caso del modelo 1, las predicciones se mantienen relativamente estables alrededor de los 23 mil puntos, sin lograr capturar la variabilidad observada en el PIB real. Esto evidencia limitaciones del modelo para adaptarse a fluctuaciones pronunciadas en el periodo de evaluación.

En el modelo 2, estimado en forma logarítmica, presenta desviaciones aún mayores en varios trimestres. Si bien la transformación logarítmica puede mejorar la estabilidad en ciertos contextos, en este caso no se traduce en una mejora de la precisión predictiva, mostrando errores considerables frente a los valores reales. Por su parte, el modelo 3, que incorpora una variable dummy para capturar efectos estructurales, exhibe predicciones muy similares al modelo 1, pero con errores ligeramente menores en la mayoría de los trimestres. Esto sugiere que la inclusión del componente estructural aporta cierta mejora marginal en la capacidad predictiva.

Al comparar el rendimiento mediante métricas de error, se observa que el modelo 3 presenta el mejor desempeño predictivo, al registrar los valores más bajos tanto en RMSE (5455.37) como en MAE (5420.39). El modelo 1 muestra un desempeño muy cercano, con diferencias marginales respecto al modelo 3, lo que indica que ambos modelos poseen una capacidad predictiva similar. En contraste,

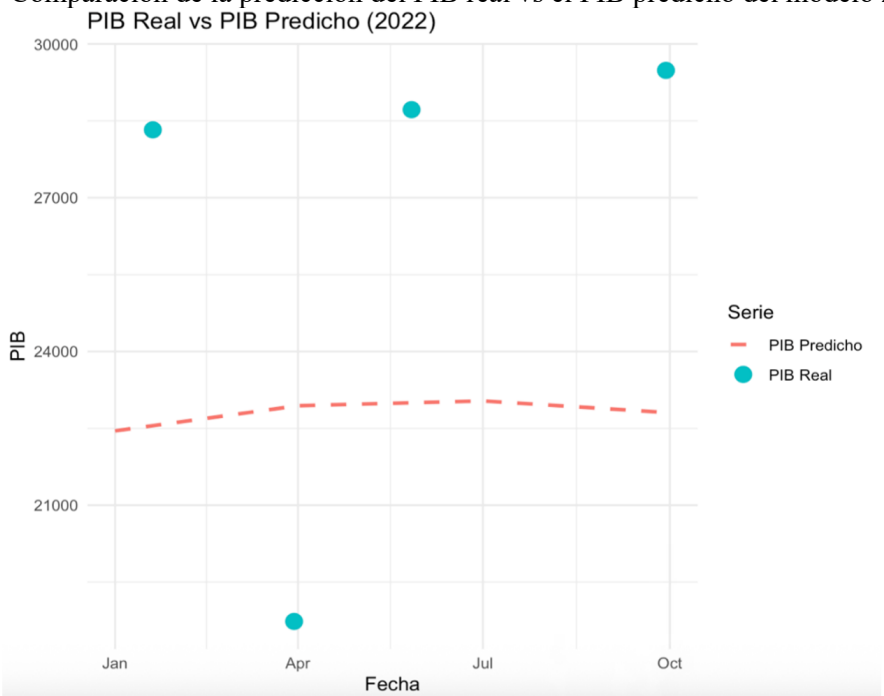
el modelo 2 reporta los mayores errores (RMSE de 5710.15 y MAE de 5645.43), evidenciando una menor precisión en sus estimaciones. Estos resultados sugieren que la inclusión de la variable dummy aporta una mejora leve en la precisión del modelo. Dichos valores se detallan a continuación.

Figura 3. Comparación de la predicción del PIB real vs el PIB predicho del modelo 1.



Fuente: elaboración propia

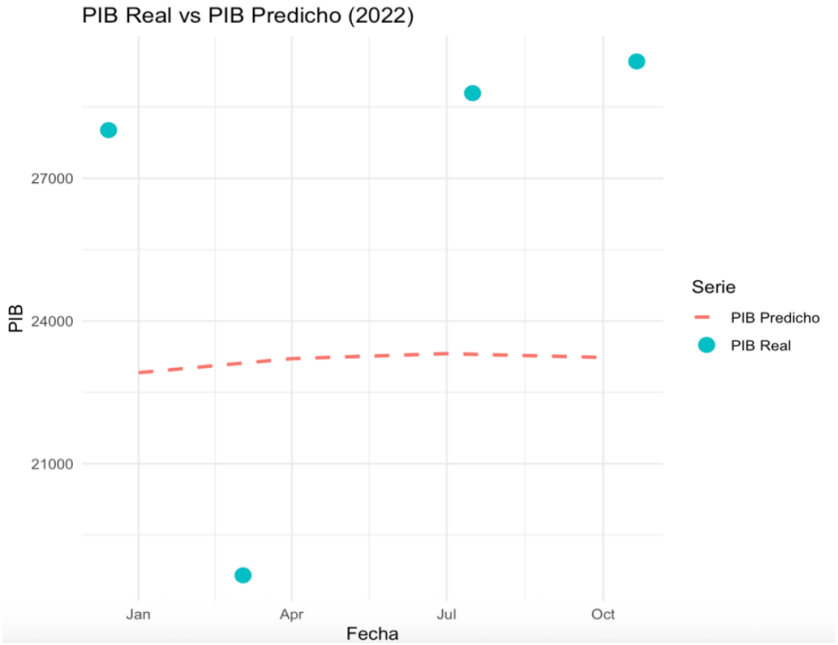
Figura 4. Comparación de la predicción del PIB real vs el PIB predicho del modelo 2



Fuente: elaboración propia.

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026 (Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

Figura 5. Comparación de la predicción del PIB real vs el PIB predicho del modelo 3.



Fuente: elaboración propia.

Modelos de rezagos distribuidos

Con el objetivo de capturar de manera más precisa la dinámica temporal del PIB y los posibles efectos retardados de las variables macroeconómicas, se incorporó un modelo de regresión lineal múltiple con rezagos distribuidos, lo que permite evaluar no solo el impacto de los factores explicativos, sino también su influencia a lo largo de varios periodos anteriores, estimándose diez variables explicativas sobre el PIB trimestral.

$$pibc = \alpha + \beta_0 * inflación_t + \beta_1 * inflación_{t-1} + \beta_2 * inflación_{t-2} + \beta_3 * inflación_{t-3} + \beta_4 * inflación_{t-4} + \gamma_0 * tiarv_t + \gamma_1 * tiarv_{t-1} + \gamma_2 * tiarv_{t-2} + \gamma_3 * tiarv_{t-3} + \gamma_4 * tiarv_{t-4} + \mu_t \quad (7)$$

A su vez con el objetivo de identificar el modelo más preciso y confiable, se llevó a cabo un proceso de evaluación exhaustiva, en el cual se iteraron todas las posibles combinaciones de las variables independientes incluidas en el modelo 4. Como resultado, se determinó que la mejor combinación de variables corresponde al modelo 5, que se puede observar a continuación.

$$pibc = \alpha + \beta_0 * inflación_t + \gamma_2 * tiarv_{t-2} + \mu_t \quad (8)$$

De igual manera se pueden visualizar los resultados de la aplicación de los modelos en la siguiente tabla.

Tabla 5. Coeficientes de los modelos.

Variables	Modelo 4	Modelo 5
Intercepto	41363***	74572***
	(18399)	11308

Inflación	-41733 (208599)	-446900 184907*
tiarv	3288 (3536)	- -
Inflación_lag 1	-224321 (186678)	- -
Inflación_lag 2	-297372 (201327)	- -
Inflación_lag 3	-302716 (200094)	- -
Inflación_lag 4	-367856 (212317)	- -
tiarv_lag1	-2002 (4786)	- -
tiarv_lag2	50 (4931)	-4740 1070***
tiarv_lag3	-1706 (4783)	- -
tiarv_lag4	-1060 (2739)	- -
R2	0.56	0.42

Nota: Nivel de significancia al 1% (***p<0.01), 5% (**p<0.05), 10% (*p<0.1). Los errores estándar robustos se muestran en paréntesis. La variable dependiente corresponde al PIB.

El modelo 4 constituye una especificación amplia con hasta cuatro rezagos de las variables explicativas, acorde con la periodicidad trimestral de los datos. Sin embargo, los resultados muestran que, a excepción del intercepto, ninguno de los coeficientes asociados a inflación, tasa de interés ni sus rezagos presenta significancia estadística.

Este resultado puede explicarse por la presencia de multicolinealidad, lo cual se confirma en la Tabla 7. En particular, los factores de inflación de la varianza (VIF) para los rezagos de la tasa de interés alcanzan valores elevados (superiores a 10 e incluso cercanos a 18), superando ampliamente los umbrales comúnmente aceptados. Esta situación genera estimadores inestables y reduce la precisión estadística de los coeficientes, aun cuando el modelo presenta un de 0.56, lo que indica una capacidad explicativa moderada en términos globales.

Por su parte, el modelo 5 presenta una especificación más parsimoniosa. En este caso, el intercepto mantiene un coeficiente positivo y estadísticamente significativo. La inflación contemporánea muestra un coeficiente negativo y significancia estadística, sugiriendo una relación inversa entre inflación y PIB, consistente con la teoría macroeconómica en contextos de presiones inflacionarias. Asimismo, la tasa de interés con dos rezagos resulta estadísticamente significativa y con signo negativo, lo que evidencia un efecto contractivo sobre el PIB con rezago temporal.

La tabla 6 confirma que en el modelo 5 los valores VIF disminuyen considerablemente (cercanos a 1), lo que indica ausencia de multicolinealidad severa. Aunque su (0.42) es menor al del modelo 4, este modelo ofrece estimaciones más estables y económicamente interpretables. En este sentido, el modelo 5 representa un mejor equilibrio entre parsimonia, consistencia teórica y solidez estadística.

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026
(Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

Tabla 6. Multicolinealidad de los modelos

Rezagos	Modelo 4.	Modelo 5.
inflación	1.71	1.12
tiarv	10.88	
inflacion_lag1	1.377	
inflacion_lag2	1.65	
inflacion_lag3	2.08	
inflación_lag4	2.93	
tiarv_lag1	18.6	
tiarv_lag2	18.11	1.12
tiarv_lag3	17.88	
tiarv_lag4	8.51	

Fuente: elaboración propia.

Mientras que la prueba de Breusch-Pagan confirmó heterocedasticidad, la presencia de este no afecta la validez del modelo ni su capacidad predictiva lo cual se puede corroborar en la Tabla 6 mediante las métricas de error de los modelos, consolidando al modelo 5 como el más acertado.

Tabla 7. Métricas de error de los modelos aplicados

	Modelo 4	Modelo 5
RMSE	5654.59	3529.86
MAE	5582.79	2257.05

Fuente: elaboración propia.

A continuación, se podrá evidenciar los valores predichos de los dos modelos de rezagos distribuidos, en la Tabla 8, así como verificar de forma visual la aproximación de cada modelo al PIB real en la figura 6.

Tabla 8. Valores predichos de cada modelo

Fecha	Trimestre	PIB Real	PIB Predicho Modelo 4	PIB Predicho Modelo 5
2022-01-01	2022I	28198	23668.04	26444.03
2022-04-01	2022II	18558	23416.60	25389.08
2022-07-01	2022III	28685	22302.20	28436.21
2022-10-01	2022IV	29607	23047.28	29801.35

Fuente: elaboración propia.

Random Forest

En función de los resultados obtenidos mediante la regresión lineal múltiple con rezagos distribuidos del modelo 5, se ajustó el modelo Random Forest para incorporar exclusivamente las variables que mostraron mayor significancia estadística y relevancia económica, la inflación actual y la tasa de interés rezagada dos periodos (tiarv_lag2). Este ajuste tuvo como propósito mejorar la

precisión predictiva del modelo y reducir la complejidad innecesaria, priorizando la eficiencia computacional y la interpretabilidad.

Este modelo se estructuró con 500 árboles de decisión y se seleccionaron dos variables aleatorias por nodo (mtry=2). La evaluación de la importancia relativa de cada predictor se realizó mediante dos métricas clave de regresión: el Incremento porcentual del error cuadrático medio (%IncMSE), el cual indica el aumento del error del modelo al permutar aleatoriamente una variable, y el Incremento en la calidad de la partición (IncNodePurity), que en el contexto de regresión mide la reducción total de la suma de los cuadrados residuales (RSS) que resulta de las particiones basadas en dicha variable.

Los resultados mostraron que tiarv_lag2 fue la variable más influyente en las predicciones, con un incremento del 45% en el error al omitirla, seguida de la inflación con un impacto del 18,5%. Matemáticamente, la predicción generada por este modelo de regresión se representa como el promedio de las predicciones de los árboles individuales representa como:

$$\hat{Y}_t = 1/500 \sum_{\beta=1}^{500} T_\beta(X_t) \quad (9)$$

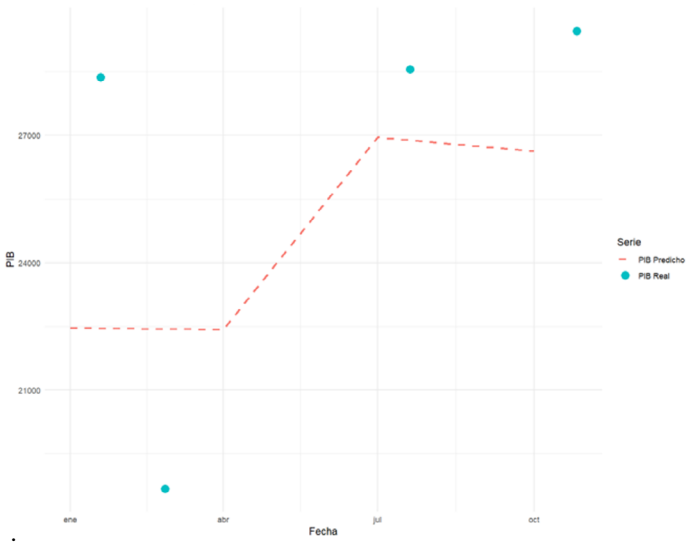
En cuanto al desempeño predictivo, el modelo presentó un RMSE de 3867.10 y un MAE de 3597.85, indicando que existe una buena capacidad de predicción general.

Tabla 9. Valores predichos del modelo Random Forest

Fecha	**Trimestre**	**PIB Real**	**PIB Predicho 6**
2022-01-01	2022I	28198	22453
2022-04-01	2022II	18558	22425
2022-07-01	2022III	28685	26955
2022-10-01	2022IV	29607	26630

Fuente: elaboración propia.

Figura 6. Comparación de la predicción del PIB real vs el PIB predicho del modelo



Fuente: elaboración propia.

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026 (Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

Long – Short Term Memory

Como parte del enfoque comparativo del presente estudio, se incorporó un modelo de red neuronal de tipo LSTM. Este modelo se seleccionó por su capacidad para capturar relaciones no lineales y dependencias secuenciales en series temporales económicas, permitiendo ajustar las predicciones del PIB en función de patrones históricos detectados en trimestres anteriores.

Para esto, se estructuraron los datos como secuencias multivariadas de longitud 4 (un año de memoria temporal), considerando como variables explicativas el PIB, la inflación y la tasa de interés. Dada la dimensión del conjunto de datos (56 observaciones), el modelo se diseñó bajo un enfoque de arquitectura parsimoniosa para prevenir el sobreajuste (overfitting).

Este modelo consta de una capa única LSTM compuesta por 50 unidades ocultas. Para compensar la complejidad de la red frente al tamaño de la muestra, se incorporaron mecanismos de regularización técnica, incluyendo una penalización de pesos (L2) y una tasa de aprendizaje controlada. La salida se procesa a través de una capa densa que transforma la información en una única predicción para el valor del PIB del siguiente trimestre. El arreglo tridimensional de entrada se define como (n_observaciones, 4 trimestres, 3 variables).

Por cada época, se calcularon los gradientes de error y se actualizó los pesos de la red con el fin de reducir la función pérdida. El valor del error se monitoreó durante todo el proceso y mostró una tendencia decreciente, indicando que el modelo aprende adecuadamente.

Como resultado del entrenamiento, se obtuvo un modelo capaz de predecir el comportamiento del PIB en función de los patrones aprendidos a lo largo del período del 2008-2021. En la siguiente tabla se presentan los valores reales del PIB frente a los valores predichos del modelo LSTM para los cuatro trimestres del año 2022.

Tabla 10. Comparación del PIB real vs PIB predicho

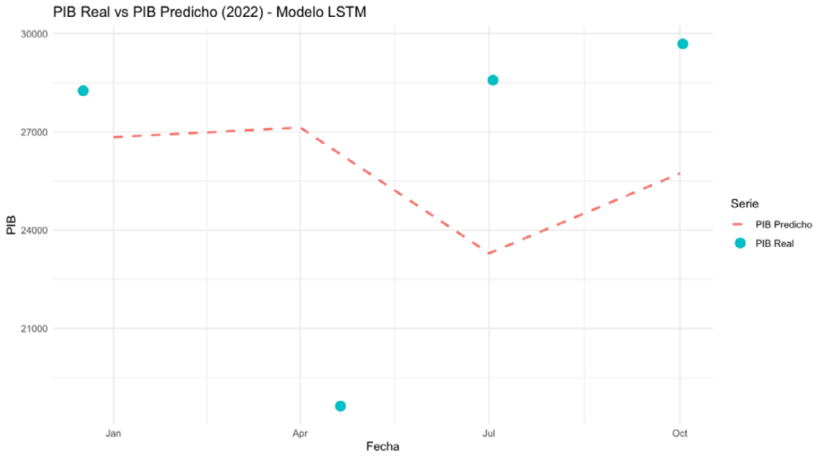
Fecha	Trimestre	PIB Real	PIB Predicho 7
2022-01-01	2022I	28198	26837.55
2022-04-01	2022II	18558	27134.57
2022-07-01	2022III	28685	23294.36
2022-10-01	2022IV	29607	25732.34

Fuente: elaboración propia.

A continuación, se presenta la representación gráfica de los resultados, lo que permite visualizar de forma más clara la precisión del modelo al comparar el comportamiento del PIB real y PIB predicho del año 2022.

Se puede observar que el modelo tiende a subestimar o sobreestimar con ciertas variabilidades, especialmente en el segundo trimestre, en el cual el valor del PIB real fue considerablemente inferior al valor predicho por el modelo, indicando una sobreestimación significativa, y para el primer y cuarto trimestre, las predicciones son más cercanas al valor real.

Figura 7. Comparación de la predicción del PIB real vs el PIB predicho del modelo



Fuente: elaboración propia.

Para culminar podemos observar en la Tabla 11 que el modelo no superó en precisión a los otros modelos, sin embargo, logró detectar la dinámica no lineal de la serie temporal, lo que demuestra la capacidad de las redes neuronales para modelar relaciones secuenciales complejas.

Tabla 11. Métricas de error del modelo LSTM.

	Modelo 7
RMSE	5465.35
MAE	4800.58

Fuente: elaboración propia.

En el campo del modelado económicos y el aprendizaje automático, las métricas de evaluación desempeñan un papel fundamental para determinar la efectividad de los modelos predictivos. Entre las medidas más utilizadas se encuentran la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE), las cuales permiten cuantificar las diferencias entre los valores reales y los valores estimados por cada modelo. Mientras el RMSE penaliza con mayor intensidad los errores de gran magnitud, el MAE mide el promedio absoluto de las desviaciones. En ambos casos, menores valores indican una mayor precisión predictiva.

Una vez estimados todos los modelos y evaluadas sus métricas de error individuales, se procedió a realizar una comparación conjunta de los valores predichos, con el fin de concluir cuál modelo reproduce con mayor fidelidad el comportamiento real del PIB. La siguiente tabla consolida los resultados de predicción de los modelos aplicados, permitiendo una revisión directa de su desempeño frente a los datos observados.

Se observa que el modelo 1, correspondiente a la regresión lineal múltiple, presentó un ajuste razonable en los trimestres I y IV, mientras que en los trimestres intermedios sus predicciones mostraron una ligera desviación respecto al valor real, aunque dentro de márgenes aceptables. En contraste, el modelo 2 que incorpora transformaciones logarítmicas, sobreestimó el PIB en todos los trimestres, evidenciando que esta transformación no aportó una mejora sustancial

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026 (Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

a la precisión predictiva, la que se refleja también en sus métricas de error más elevadas. Por su parte, el modelo 3, que incluyó una variable dummy para capturar posibles efectos de la pandemia, ofreció predicciones similares al modelo 1.

Tabla 12. PIB real vs PIB predichos de los modelos aplicados

	PIB Real	PIB Predicho 1	PIB Predicho 2	PIB Predicho 3	PIB Predicho 4	PIB Predicho 5	PIB Predicho 6	PIB Predicho 7
2022I	28198	27104	29307	27392.43	23668.04	26444.03	24816.50	26837.55
2022II	18558	27014	30793	27283.64	23416.60	25389.08	23946.88	27134.57
2022III	28685	29313	33028	29707.96	22302.20	28436.21	24223.03	23294.36
2022IV	29607	29460	32995	29867.05	23047.28	29801.35	24482.32	25732.34

Fuente: elaboración propia.

En cuanto al modelo 4, que incorporó rezagos distribuidos, se evidenciaron problemas importantes, especialmente en el cuarto trimestre donde la predicción fue inusualmente elevada. Como resultado, se desarrolló el modelo 5 que simplificó la estructura del modelo anterior y corrigió los problemas detectados, logrando predicciones más ajustadas del PIB real en todos los trimestres, especialmente en el tercer y cuarto, lo que se consolidó con sus métricas de error más favorables.

En el ámbito de los modelos no lineales, el modelo 6 basado en Random Forest mostró un comportamiento estable, con predicciones moderadamente cercanas al PIB real, aunque tendió a subestimar los valores en los trimestres intermedios. Finalmente, el modelo 7, implementado mediante redes neuronales LSTM, capturó de manera efectiva los patrones secuenciales de las variables, generando predicciones ajustadas en los trimestres I y IV con leves desviaciones en los trimestres restantes, manteniendo una coherencia general en sus resultados.

En síntesis, el modelo 5, resultado de un proceso de optimización sobre el modelo de rezagos distribuidos, demostró ser el más acertado dentro del escenario caracterizado por alta volatilidad económica y presencia de choques exógenos durante el periodo de prueba 2022. Al simplificar la estructura inicial del modelo 4 y conservar únicamente la inflación contemporánea y el rezago de la tasa de interés correspondiente a dos trimestres previos, se obtuvo una especificación parsimoniosa que redujo significativamente los problemas de multicolinealidad y mejoró la capacidad predictiva. Este modelo presentó los menores niveles de error dentro de dicho contexto y logró aproximarse de manera consistente al comportamiento observado del PIB real.

No obstante, al evaluar los modelos en un entorno macroeconómico más estable, utilizando el periodo 2008-2018 como entrenamiento y 2019 como prueba, el modelo LSTM presentó mejor desempeño predictivo. Esto evidencia que la efectividad de los modelos depende de las condiciones económicas analizadas, ya que los modelos lineales mostraron mayor robustez ante escenarios de alteraciones estructurales, mientras que los modelos de aprendizaje automático lograron mejores resultados en contextos de mayor estabilidad.

Evaluación complementaria de modelos predictivos en un entorno económico estable (2008-2019)

Con el fin de realizar una validación adicional y analizar el desempeño de los modelos fuera del contexto de alta incertidumbre observado entre (2020 – 2022), se aplicaron los mismos modelos predictivos utilizando un periodo alternativo (2008 – 2018) como entrenamiento y 2019 como prueba. Este escenario se caracteriza por una mayor estabilidad macroeconómica sin eventos exógenos externos, lo cual permitió observar el comportamiento de los modelos bajo condiciones más normales del ciclo económico.

En este contexto, el modelo LSTM demostró un rendimiento significativamente superior, logrando el menor error cuadrático y confirmando su eficacia en la predicción del PIB cuando las condiciones económicas no presentan shocks inesperados.

A continuación, se presenta la tabla que resume los resultados obtenidos.

Tabla 13. PIB real vs PIB predichos de los modelos alternos

Fecha	Trimestre	PIB Real	PIB Predicho 1	PIB Predicho 2	PIB Predicho 3	PIB Predicho 4	PIB Predicho 5	PIB Predicho 6	PIB Predicho 7
1/1/2019	2019I	28198	26795	27923	26795	26429	26964.18	25258	27036
1/4/2019	2019II	18558	25689	26226	25689	28512	27467.56	25649	27077
1/7/2019	2019III	28685	25588	26016	25588	29042	26984.94	25289	27082
1/10/2019	2019IV	29607	25924	26450	25924	30197	26794.37	26076	27119

Fuente: elaboración propia.

Tabla 14. Métricas de error de los modelos predictivos en un entorno económico estable (2008-2019)

Modelos	MAE	RMSE
Modelo 1.	1027.45	1158.98
Modelo 2.	877.06	982.13
Modelo 3.	1027.45	1158.98
Modelo 4.	1761.12	2001.20
Modelo 5.	217.60	258.59
Modelo 6.	1458.67	1487.70
Modelo 7.	**51.91**	**66.21.**

Fuente: elaboración propia

Los resultados muestran que el modelo 7 fue el más preciso en un entorno económico estable, registrando los menores niveles de error entre todos los modelos evaluados. Su capacidad para aprender patrones secuenciales le permitió predecir con gran exactitud el comportamiento del PIB, consolidándolo como el modelo más eficaz para contextos sin perturbaciones externas.

Limitaciones

El presente estudio presenta algunas limitaciones que deben ser consideradas. En primer lugar, el tamaño de la muestra es relativamente reducido debido a la disponibilidad de datos trimestrales para el periodo analizado, lo que restringe la

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026
(Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

capacidad de entrenamiento de modelos complejos como LSTM.

Asimismo, las variables ya habían sido previamente identificadas, mediante análisis exploratorios y criterios econométricos como el conjunto de indicadores con mayor relevancia predictiva y mayores niveles de correlación respecto al PIB dentro de las variables macroeconómicas analizadas. Por ello, no se incorporaron variables adicionales, priorizando modelos parsimoniosos con el fin de reducir posibles problemas de sobreajuste y multicolinealidad.

Adicionalmente, con el propósito de reducir riesgos de sobreajuste, se utilizaron particiones temporales separadas para entrenamiento y prueba, respetando la estructura cronológica de las series de tiempo. También se priorizaron modelos parsimoniosos y configuraciones controladas de hiperparámetros, especialmente en los modelos de aprendizaje automático, evitando estructuras excesivamente complejas en relación con el tamaño de la muestra disponible.

Por otro lado, debido a las restricciones de información y al tamaño muestral, no fue posible implementar esquemas más complejos de validación cruzada ni procesos extensivos de optimización de hiperparámetros. Asimismo, la presencia de eventos extraordinarios, como la pandemia de COVID-19 y el paro nacional, introdujo cambios estructurales que pueden afectar la estabilidad de los modelos y limitar la generalización de los resultados.

Por último, una limitación adicional radica en la escasez de investigaciones previas que integren simultáneamente modelos econométricos tradicionales y técnicas de aprendizaje automático aplicadas a economías dolarizadas como la ecuatoriana, lo que restringe la posibilidad de realizar comparaciones directas con evidencia empírica equivalente. Sin embargo, esta situación también evidencia el carácter novedoso del enfoque propuesto y su potencial contribución a la literatura sobre predicción macroeconómica en contextos emergentes.

Discusión y conclusiones

El presente estudio tuvo como objetivo evaluar la capacidad predictiva de distintos modelos aplicados al PIB real del Ecuador, combinando enfoques econométricos tradicionales y técnicas de aprendizaje automático. A diferencia de investigaciones centradas en la identificación de relaciones causales, este trabajo priorizó la comparación del desempeño predictivo mediante métricas de error como el RMSE y el MAE.

Los resultados evidenciaron que los modelos de regresión lineal múltiple, estimados bajo mínimos cuadrados ordinarios y corregidos mediante mínimos cuadrados generalizados (GLS) con estructura AR (1) ante la presencia de autocorrelación, presentaron un desempeño predictivo competitivo. En particular, el Modelo 3 registró los menores valores de RMSE y MAE entre las especificaciones de regresión lineal múltiple consideradas, lo que indica un mejor desempeño relativo dentro de este grupo de modelos.

Esto sugiere que la incorporación de componentes estructurales, como variables dummy para eventos atípicos, puede contribuir a mejorar la precisión de las predicciones en presencia de choques económicos.

En contextos de alta incertidumbre y presencia de eventos exógenos, como la pandemia y el paro nacional, los modelos econométricos mostraron una capacidad predictiva robusta. En particular, el Modelo 5, basado en una especificación optimizada de rezagos distribuidos, presentó el mejor desempeño en el escenario de alta volatilidad correspondiente al periodo de prueba 2022. Esto indica que, bajo escenarios de elevada volatilidad, modelos más parsimoniosos y con estructuras adecuadamente especificadas pueden ofrecer resultados estables y consistentes.

Por otro lado, al evaluar los modelos en un entorno relativamente más estable, el modelo LSTM mostró un desempeño superior en términos de reducción del error de predicción. Este resultado pone de manifiesto que los modelos basados en aprendizaje automático pueden capturar de mejor manera patrones no lineales y dinámicas intertemporales cuando la estructura de los datos es más estable.

En conjunto, los hallazgos confirman que no existe un modelo universalmente superior, sino que su desempeño depende del contexto económico y de la naturaleza de los datos. La complementariedad entre econometría tradicional y aprendizaje automático constituye, por tanto, una estrategia valiosa para la predicción económica.

Finalmente, a pesar de las limitaciones señaladas, este enfoque resulta especialmente relevante para economías dolarizadas, como la ecuatoriana, donde las herramientas de política monetaria son limitadas.

La aplicación de modelos predictivos confiables puede contribuir a mejorar la planificación fiscal, la evaluación de riesgos macroeconómicos y el diseño de políticas públicas. Asimismo, países con regímenes monetarios similares, como Panamá o El Salvador, podrían beneficiarse de metodologías comparables para anticipar el comportamiento de variables macroeconómicas clave.

Declaración de contribución de autoría CRediT

Estéfany A. Orbe-Rosario: Conceptualización, análisis formal, metodología, validación y redacción del borrador original.

Tatiana L. Palacios-Sinchi: Análisis formal, Investigación, metodología, redacción del borrador original.

Pablo Gaibor Costta: Análisis formal, Investigación, metodología, redacción del borrador original.

Declaración de conflictos de interés

Los autores declaran no tener ningún conflicto de intereses.

Referencias

1. Breiman, L. (2001). *Random Forests.* Recuperado el 13 de junio de 2025, de https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

KAIRÓS, revista de ciencias económicas, jurídicas y administrativas, 9(17), pp. 93-117. Segundo Semestre de 2026 (Ecuador). ISSN 2631-2743. DOI: https:/10.37135/kai.03.17.05

2. Brownlee, J. (2020). *Machine Learning Mastery.* (M. L. Mastery, Ed.) Recuperado el 14 de junio de 2025, de Machine Learning Mastery: https://machinelearningmastery.com/machine-learning-for-time-series-forecasting/

3 DataScientest. (20 de mayo de 2024). *DataScientest*. Recuperado el 6 de junio de 2025, de DataScientest: https://datascientest.com/es/memoria-a-largo-plazo-a-corto-plazo-lstm#:~:text=Fueron%20propuestas%20por%20J%C3%BCrgen%20Schmidhuber,espec%C3%ADficamente%20para%20superar%20este%20problema.

4. Espino C, Martínez X, Daradoumis A. *Análisis Predictivo: Técnicas y Modelos Utilizados y Aplicaciones del mismo—Herramientas Open Source que permiten su uso.* Trabajo de fin de grado en Ingeniería Informática en la Universitat Oberta de Catalunya. 2017. https://api.core.ac.uk/oai/oai:openaccess.uoc.edu:10609/59565

5. Gujarati, D., & Porter, D. (2010). *Econometría.* México. Recuperado el 6 de junio de 2025

6. Hyndman, R., & Athanasopoulos, G. (31 de mayo de 2021). *Otexts*. Recuperado el 14 de junio de 2025, de Otexts: https://otexts.com/fpp3/

7. Mejía García, D. D., & Acosta Pérez, B. R. (2023). Avances tecnológicos modernos y sus implicaciones en el pensamiento social. *AULA Revista de Humanidades y Ciencias Sociales, 65*(2), 29-37. https://doi.org/10.33413/aulahcs.2019.65i2.118

8 Montero, R. (marzo de 2016). *Modelos de regresión lineal multiple.* Recuperado el 8 de junio de 2025, de Universidad de Granada: https://www.ugr.es/~montero/matematicas/regresion_lineal.pdf