



Maestría en Inteligencia Artificial Aplicada

MAESTRÍA EN INTELIGENCIA ARTIFICIAL APLICADA

Trabajo previo a la obtención de título de
Magister en Inteligencia Artificial Aplicada

AUTORES:

Christian Alejandro Bonilla Pilataxi

Kevin Andres Solis Nuñez

Cristian Alejandro Veloz Armas

TUTORES:

Marcela Azucena Venegas Espinosa

Karla Estefanía Mora Cajas

Pronóstico multi-horizonte del ITCER en Ecuador mediante aprendizaje profundo con variables macroeconómicas.

Certificación de autoría

Nosotros, **Christian Alejandro Bonilla Pilataxi, Kevin Andrés Solís Núñez, Cristian Alejandro Veloz Armas**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

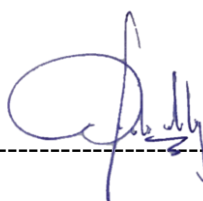
Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Christian Alejandro Bonilla Pilataxi



Kevin Andrés Solís Núñez



Cristian Alejandro Veloz Armas

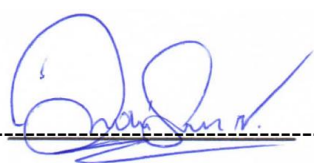
Autorización de Derechos de Propiedad Intelectual

Nosotros, **Christian Alejandro Bonilla Pilataxi, Kevin Andrés Solís Núñez, Cristian Alejandro Veloz Armas**, en calidad de autores del trabajo de investigación titulado **Pronóstico multi-horizonte del ITCER en Ecuador mediante aprendizaje profundo con variables macroeconómicas**, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

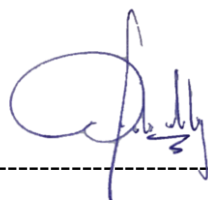
D. M. Quito, (agosto 2026)



Christian Alejandro Bonilla Pilataxi



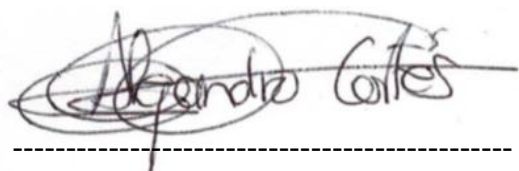
Kevin Andrés Solís Núñez



Cristian Alejandro Veloz Armas

Aprobación de dirección y coordinación del programa

Nosotros, **Alejandro Cortés López** y **Karla Estefanía Mora Cajas**, declaramos que: **Christian Alejandro Bonilla Pilataxi**, **Kevin Andrés Solís Núñez** y **Cristian Alejandro Veloz Armas** son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director de la
Maestría en Inteligencia Artificial Aplicada



Karla Estefanía Mora Cajas
Coordinadora de la
Maestría en Inteligencia Artificial Aplicada

DEDICATORIA

Dedicamos este trabajo, en primer lugar, a Dios, por darnos la fortaleza, la perseverancia y la sabiduría necesarias para culminar esta etapa de nuestra formación profesional.

A nuestras familias, por su apoyo incondicional, por acompañarnos durante este proceso y por ser una fuente constante de motivación para seguir adelante, incluso en los momentos de mayor dificultad.

A todas las personas que, de una u otra manera, contribuyeron a que este proyecto fuera posible, brindándonos su apoyo, sus conocimientos y sus palabras de aliento.

Finalmente, dedicamos este logro a nosotros mismos, por el esfuerzo, la disciplina y la constancia que implicó alcanzar esta meta y culminar la Maestría en Inteligencia Artificial Aplicada de la Universidad Internacional del Ecuador (UIDE).

AGRADECIMIENTOS

Expresamos nuestro más sincero agradecimiento a Dios, por brindarnos la fortaleza, perseverancia y sabiduría necesarias para culminar esta etapa de nuestra formación académica y afrontar los retos que se presentaron durante el desarrollo de esta investigación.

A nuestras familias, por su apoyo incondicional, comprensión y confianza. Su respaldo y motivación fueron fundamentales para mantener nuestro compromiso y constancia a lo largo de este proceso.

A la Universidad Internacional del Ecuador (UIDE) y a los docentes de la Maestría en Inteligencia Artificial Aplicada, por los conocimientos, experiencias y herramientas proporcionadas durante nuestra formación. Los aprendizajes adquiridos fueron fundamentales para fortalecer nuestras capacidades profesionales y llevar a cabo el presente trabajo.

RESUMEN

El presente trabajo desarrolla y evalúa un esquema de pronóstico multi-horizonte para el Índice de Tipo de Cambio Efectivo Real (ITCER) del Ecuador, utilizando información histórica del índice y variables macroeconómicas relacionadas.

La metodología incluyó modelos de referencia y modelos de aprendizaje automático y aprendizaje profundo. Se implementaron Naive Persistence, Seasonal Naive, ETS, SARIMAX, LightGBM, LSTM multi-horizonte, Temporal Fusion Transformer (TFT) y Time-series Dense Encoder (TiDE). La evaluación se realizó mediante métricas como MAE, RMSE, MAPE, sMAPE y R^2 , manteniendo una partición temporal de los datos para evitar el uso de información futura durante el entrenamiento y la prueba. Para los modelos multivariados se incorporaron variables como IPC de Ecuador, IPC de Estados Unidos, USD NEER, WTI y riesgo país EMBI.

En conjunto, los resultados muestran que una mayor complejidad del modelo no garantiza un mejor pronóstico para el ITCER. Los modelos más sencillos o de complejidad intermedia obtuvieron resultados competitivos, mientras que el desempeño de los modelos de aprendizaje profundo estuvo condicionado por el tamaño reducido de la muestra. El análisis de interpretabilidad permitió, además, identificar la relevancia de variables como IPC de Estados Unidos, EMBI e IPC de Ecuador dentro del modelo TFT, aunque estos resultados se interpretan como importancia predictiva y no como relaciones causales.

Palabras clave:

ITCER, pronóstico multi-horizonte, aprendizaje automático, aprendizaje profundo, variables macroeconómicas, series temporales.

ABSTRACT

This study develops and evaluates a multi-horizon forecasting framework for Ecuador's Real Effective Exchange Rate Index (REER), using historical information from the index and related macroeconomic variables.

The methodology included benchmark models, as well as machine learning and deep learning approaches. Naive Persistence, Seasonal Naive, ETS, SARIMAX, LightGBM, multi-horizon LSTM, Temporal Fusion Transformer (TFT), and Time-series Dense Encoder (TiDE) models were implemented. Model performance was evaluated using MAE, RMSE, MAPE, sMAPE, and R^2 metrics, with a temporal data partition to prevent the use of future information during training and testing. The multivariate models incorporated variables such as Ecuador's Consumer Price Index (CPI), the United States Consumer Price Index (CPI), USD NEER, WTI oil prices, and EMBI country risk.

Overall, the results show that greater model complexity does not necessarily guarantee better forecasting performance for the REER. Simpler or moderately complex models achieved competitive results, while the performance of deep learning models was conditioned by the limited sample size. The interpretability analysis also identified variables such as the United States CPI, EMBI, and Ecuador's CPI as relevant within the TFT model. These results are interpreted as measures of predictive importance rather than causal relationships.

Keywords:

REER, multi-horizon forecasting, machine learning, deep learning, macroeconomic variables, time series

Índice de contenido

ÍNDICE DE CONTENIDO	viii
ÍNDICE DE FIGURAS.....	xiii
ÍNDICE DE TABLAS	xv
1. INTRODUCCIÓN.....	1
DEFINICIÓN DEL PROYECTO.....	1
JUSTIFICACIÓN E IMPORTANCIA DEL TRABAJO DE INVESTIGACIÓN.....	3
ALCANCE.....	4
OBJETIVOS	6
1.1.1. <i>Objetivo General</i>	6
1.1.2. <i>Hipótesis</i>	6
1.1.3. <i>Objetivos Específicos</i>	6
2. REVISIÓN DE LITERATURA.....	9
2.1 ESTADO DEL ARTE.....	9
2.2 MARCO TEÓRICO	21
2.2.1 <i>Procesos estocásticos</i>	21
2.2.2 <i>Procesos estocásticos estacionarios</i>	21
2.2.3 <i>Series Temporales</i>	22
2.2.3.1 <i>Proceso de ruido blanco</i>	24
2.2.3.2 <i>Componentes de una serie temporal</i>	24
2.2.3.3 <i>Dependencia temporal</i>	25
2.2.4 <i>Pronóstico de series temporales</i>	27
2.2.4.1 <i>Series univariadas y multivariadas</i>	27
2.2.4.2 <i>Ventanas temporales</i>	27
2.2.4.3 <i>Pronóstico multi-horizonte</i>	28
2.2.4.4 <i>Pronóstico de un paso</i>	29
2.2.4.5 <i>Pronóstico de múltiples pasos</i>	29
2.2.4.6 <i>Estrategia recursiva</i>	29

2.2.4.7 Estrategia directa.....	30
2.2.4.8 Estrategia de salida múltiple.....	30
2.2.4.9 Construcción de muestras multi-horizonte	31
2.2.4.10 Evaluación por horizonte.....	31
2.2.4.11 Modelo ingenuo	32
2.2.4.12 Modelo Seasonal Naive (Ingenuo Estacional)	32
2.2.5 <i>Índice de Tipo de Cambio Efectivo Real (ITCER)</i>	33
2.2.5.1 Definición del ITCER	33
2.2.5.2 Cálculo general.....	35
2.2.5.3 ITCER en una economía dolarizada	36
2.2.5.4 Factores macroeconómicos relacionados con el ITCER	37
2.2.6 <i>Modelos estadísticos lineales y econométricos de series temporales</i>	39
2.2.6.1 Modelos Autorregresivos Integrados de Media Móvil (ARIMA).....	39
2.2.6.2 El componente Autorregresivo (AR-p) y operadores de rezago	39
2.2.6.3 El componente de Media Móvil (MA-q) y choques estocásticos.....	40
2.2.6.4 El proceso de Integración (I-d) y eliminación de no estacionariedad	41
2.2.6.5 Metodología de Box-Jenkins (Identificación, estimación y diagnóstico de residuos)	41
2.2.6.6 Modelos SARIMAX con componentes estacionales y variables exógenas	42
2.2.6.7 Extensión estacional (SARIMA): Especificación matemática de los órdenes (P,D,Q)s	42
2.2.6.8 Incorporación de covariables externas (Componente X).....	43
2.2.7 <i>Aprendizaje automático aplicado al pronóstico económico</i>	45
2.2.8 <i>Aprendizaje supervisado</i>	46
2.2.8.1 Validación temporal	47
2.2.8.2 Modelos basados en árboles de decisión	48
2.2.8.3 Gradient Boosting.....	50
2.2.7 <i>LightGBM</i>	52
2.2.7.5 <i>Definición</i>	52
2.2.9.2 <i>Gradient Boosting y aprendizaje secuencial</i>	53
2.2.9.3 <i>Construcción de árboles mediante histogramas</i>	55
2.2.9.4 <i>Crecimiento por hojas</i>	56
2.2.8 <i>Aprendizaje profundo para series temporales</i>	57
2.2.8.5 <i>Redes neuronales artificiales</i>	57

2.2.8.6	<i>Redes neuronales recurrentes (RNN)</i>	59
2.2.8.7	<i>Long Short-Term Memory (LSTM)</i>	60
2.2.8.8	<i>LSTM Multi-horizonte</i>	61
2.2.8.9	<i>Temporal Fusion Transformer (TFT)</i>	62
2.2.8.10	<i>Interpretabilidad en modelos TFT</i>	63
2.2.8.11	<i>Interpretabilidad mediante la selección de variables</i>	63
2.2.8.12	<i>Interpretabilidad mediante Self-Attention</i>	64
2.2.8.13	<i>TiDE / Time-series Dense Encoder</i>	65
2.2.8.14	<i>Arquitectura del Modelo TiDE</i>	66
3	DESARROLLO	71
3.1	DESCRIPCIÓN LAS VARIABLES	71
3.1.1	<i>Fuentes de datos</i>	71
3.1.2	<i>Definición de variables del modelo</i>	72
3.1.3	<i>Variables objetivo</i>	72
3.1.4	<i>Variables exógenas</i>	72
3.1.5	<i>Variables calendario</i>	75
4	DESARROLLO DEL PROYECTO	76
4.1	CONSTRUCCIÓN Y PREPROCESAMIENTO DE LA BASE DE DATOS	76
	DISEÑO EXPERIMENTAL	76
4.2	MODELOS BASE PARA COMPARACIÓN	77
4.2.1	ESTIMACIÓN CON MODELO NAIVE PERSISTENCE	77
4.2.2	Datos utilizados y partición temporal	78
4.2.3	<i>Estimación con modelo Seasonal Naive</i>	82
4.2.3.1	Evaluación del modelo	83
4.2.3.2	Influencia de los valores atípicos	84
4.2.3.3	Comportamiento del modelo en un horizonte de largo plazo	84
4.2.4	<i>Estimación modelo ETS</i>	85
4.2.4.1	Implementación de modelos clásicos de suavizamiento exponencial	85
4.2.4.2	Entrenamiento del modelo ETS	87

4.2.4.3	Evaluación del modelo ETS.....	88
4.2.4.4	Comportamiento del modelo ETS en un horizonte de largo plazo	89
4.2.5	<i>Estimación modelo SARIMAX</i>	89
4.2.5.1	Análisis de estacionariedad	89
4.2.4.2	Identificación y selección del modelo ARIMA.....	90
4.2.6	<i>LightGBM directo multi-horizonte</i>	100
4.2.6.1	Datos utilizados y partición temporal	101
4.2.6.2	Variables utilizadas.....	102
4.2.6.3	Configuración del modelo.....	102
4.2.6.4	Comparación por número de arboles	103
4.2.6.5	Mejores configuraciones por horizonte ^[OBJ]	104
4.2.6.6	Predicciones	104
4.2.6.7	Comparación con Naive Persistence	106
4.2.6.8	Resultados	107
4.2.7	<i>Estimación modelo LSTM multi-horizonte</i>	108
4.2.7.1	Evaluación del desempeño del modelo	113
4.2.8	<i>Estimación Modelo TFT (Temporal Fusion Transformer)</i>	114
4.2.8.1	Datos utilizados y partición temporal	116
4.2.8.2	Variables utilizadas ^[OBJ]	117
4.2.8.3	Adecuación de enfoques univariados y multivariados ^[OBJ]	118
4.2.8.4	Configuración del modelo.....	118
4.2.8.5	Parámetros entrenables	120
4.2.8.6	Funcionamiento general de la arquitectura.....	120
4.2.8.7	Interpretabilidad del TFT	121
4.2.8.8	Importancia de variables históricas	121
4.2.8.9	Importancia de variables futuras conocidas	123
4.2.8.10	Atención temporal	124
4.2.8.11	Experimento de selección de variables exógenas (ablación)	125
4.2.8.12	Predicciones del modelo seleccionado	126
4.2.8.13	Comparación con otros modelos del proyecto	127
4.2.8.14	Pronostico consolidado h1, h3 y h6	128
4.2.9	<i>Estimación Modelo TiDE</i>	129
4.2.9.1	Variables utilizadas.....	130

4.2.9.2	Tamaño de los modelos	131
4.2.9.3	Comparación entre instancias	132
4.2.9.4	Predicciones del modelo seleccionado.....	133
5	ANÁLISIS DE RESULTADOS	136
5.1	ANÁLISIS DESCRIPTIVO DE LOS DATOS	136
5.2	RESULTADOS DEL ENTRENAMIENTO DEL MODELO.....	139
5.2	RESULTADOS DE COMPARATIVA DE LOS MODELOS UTILIZADOS	141
5.2.1	<i>Evaluación del desempeño predictivo.....</i>	<i>142</i>
5.2.2	<i>Resultados del pronóstico multi-horizonte.....</i>	<i>144</i>
5.2.3	<i>Análisis de interpretabilidad del modelo</i>	<i>145</i>
6	CONCLUSIONES Y RECOMENDACIONES	147
6.1	CONCLUSIONES	147
6.2	RECOMENDACIONES.....	149
	REFERENCIAS BIBLIOGRÁFICAS.....	150

Índice de Figuras

FIGURA 1 <i>EVOLUCIÓN MENSUAL ITCER</i>	34
FIGURA 2 <i>ARQUITECTURA DEL CODIFICADOR DE TIDE</i>	67
FIGURA 3 <i>COMPARACIÓN DE PREDICCIONES PARA MODELOS NAIVE</i>	81
FIGURA 4 <i>PRONÓSTICOS DEL MODELO DE PERSISTENCIA PARA LOS HORIZONTES H1, H3 Y H6</i>	81
FIGURA 5 <i>PRONÓSTICO DEL MODELO SEASONAL NAIVE PARA EL ITCER</i>	82
FIGURA 6 <i>RESIDUOS DEL MODELO SEASONAL NAIVE</i>	83
FIGURA 7 <i>PRONÓSTICO DEL MODELO SEASONAL NAIVE EXTENDIDO HASTA EL AÑO 2030</i>	85
FIGURA 8 <i>DESCOMPOSICIÓN DE ITCR: TENDENCIA, ESTACIONALIDAD Y RESIDUOS</i>	86
FIGURA 9 <i>EVALUACIÓN DEL MODELO ETS PARA ITCER</i>	88
FIGURA 10 <i>PRONÓSTICO DE ITCER APLICANDO ETS</i>	89
FIGURA 11 <i>ANÁLISIS ACF Y PACF DE ITCER</i>	91
FIGURA 12 <i>MODELOS SARIMA ORDENADOS POR RMSE</i>	92
FIGURA 13 <i>RESULTADOS ENTRENAMIENTO SARIMAX CON TODAS LAS VARIABLES MACROECONÓMICAS</i>	94
FIGURA 14 <i>RESULTADOS ENTRENAMIENTO SARIMAX CON VARIABLES MÁS REPRESENTATIVAS</i>	95
FIGURA 15 <i>RESULTADOS ENTRENAMIENTO SARIMAX HÍBRIDO</i>	98
FIGURA 16 <i>PRONÓSTICO SARIMAX HÍBRIDO 12 MESES PARA ITCER</i>	100
FIGURA 17 <i>COMPARACIÓN ENTRE LOS VALORES REALES Y PREDICHOS DEL MODELO LIGHTGBM PARA LOS HORIZONTES H1, H3 Y H6</i>	107
FIGURA 18 <i>SERIE HISTÓRICA DEL ITCER Y AJUSTES DEL MODELO LIGHTGBM PARA LOS HORIZONTES H1, H3 Y H6</i>	108
FIGURA 19 <i>FUNCIÓN DE PERDIDA DE MODELO LSTM 1</i>	110
FIGURA 20 <i>FUNCIÓN DE PERDIDA DE MODELO LSTM 2</i>	111
FIGURA 21 <i>FUNCIÓN DE PERDIDA DE MODELO LSTM 3</i>	111
FIGURA 22 <i>PRONÓSTICOS OBTENIDOS MEDIANTE LOS MODELOS LSTM MULTI-HORIZONTE: MODELO 1</i>	112
FIGURA 23 <i>PRONÓSTICOS OBTENIDOS MEDIANTE LOS MODELOS LSTM MULTI-HORIZONTE: MODELO 2</i>	112
FIGURA 24 <i>PRONÓSTICOS OBTENIDOS MEDIANTE LOS MODELOS LSTM MULTI-HORIZONTE: MODELO 2</i>	113
FIGURA 25 <i>ENCODER DE TFT</i>	123
FIGURA 26 <i>ENTRENAMIENTO DE TFT</i>	126
FIGURA 27 <i>PRONÓSTICO TFT</i>	128

FIGURA 28 <i>PRONÓSTICO TFT</i>	134
FIGURA 29 <i>AJUSTES DE MODELOS ENTRENADOS</i>	135
FIGURA 30 <i>ESTADÍSTICA DESCRIPTIVA DE LOS DATOS DEL ITCER DEL ECUADOR</i>	136
FIGURA 31 <i>EVOLUCIÓN TEMPORAL DEL ITCER DURANTE EL PERÍODO DE ESTUDIO</i>	137
FIGURA 32 <i>DIAGRAMA DE CAJA (BOXPLOT) DE LA DISTRIBUCIÓN DEL ITCER</i>	137
FIGURA 33 <i>CORRELACIÓN DE VARIABLES EXÓGENAS</i>	138

Índice de Tablas

TABLA 1 <i>VARIABLES EXÓGENAS DEL MODELO</i>	74
TABLA 2 <i>ELABORACIÓN PROPIA A PARTIR DEL DATASET DE MODELADO</i>	78
TABLA 3 <i>MÉTRICAS DE MODELO NAIVE POR CADA HORIZONTE</i>	79
TABLA 4 <i>MÉTRICAS DE MODELOS NAIVE PERSISTENCE</i>	79
TABLA 5 <i>ERRORES OBTENIDOS EN EL MODELO NAIVE</i>	80
TABLA 6 <i>RESULTADOS MÉTRICAS CON Y SIN OUTLIERS</i>	84
TABLA 7 <i>RESULTADOS DE LA PRUEBA AUMENTADA DE DICKEY-FULLER (ADF) PARA ITCER</i>	90
TABLA 8 <i>MÉTRICAS DE EXPERIMENTOS SARIMA</i>	91
TABLA 9 <i>COEFICIENTES DE VARIABLES MACROECONÓMICAS</i>	93
TABLA 10 <i>MÉTRICAS SARIMAX EXPERIMENTO 1</i>	95
TABLA 11 <i>MÉTRICAS SARIMAX CON VARIABLES MÁS REPRESENTATIVAS</i>	96
TABLA 12 <i>COEFICIENTES DE REZAGOS DE VARIABLES MACROECONÓMICAS</i>	96
TABLA 13 <i>MÉTRICAS SARIMAX HÍBRIDO</i>	98
TABLA 14 <i>SEGMENTACIÓN DE DATASET PARA EVALUACIÓN DEL MODELO LIGHTGBM</i>	101
TABLA 15 <i>HIPERPARÁMETROS UTILIZADOS PARA LA ESTIMACIÓN DEL MODELO LIGHTGBM</i>	102
TABLA 16 <i>MÉTRICA MAE DE PRUEBAS POR HORIZONTE</i>	103
TABLA 17 <i>MEJORES CONFIGURACIONES POR CADA HORIZONTE</i>	104
TABLA 18 <i>RESUMEN DE PREDICCIONES REALIZADAS</i>	105
TABLA 19 <i>MÉTRICAS POR CADA HORIZONTE</i>	105
TABLA 20 <i>COMPARACIÓN LIGHTGBM CON NAIVE PERSISTENCE</i>	106
TABLA 21 <i>ARQUITECTURAS EVALUADAS DEL MODELO LSTM MULTI-HORIZONTE</i>	109
TABLA 22 <i>MÉTRICAS DE DESEMPEÑO DE LOS MODELOS LSTM MULTI-HORIZONTE</i>	113
TABLA 23 <i>CONFIGURACIÓN DEL MODELO TFT</i>	119
TABLA 24 <i>PARÁMETROS ENTRENABLES DE TFT</i>	120
TABLA 25 <i>IMPORTANCIA DE VARIABLES HISTÓRICAS</i>	122
TABLA 26 <i>COMPARATIVA HORIZONTES 1, 3 Y 6 MESES</i>	123
TABLA 27 <i>ATENCIÓN TEMPORAL</i>	124
TABLA 28 <i>RANKING DEL MAE PROMEDIO GLOBAL</i>	125

TABLA 29 <i>COMPARATIVA ENTRE PREDICCIONES TFT</i>	127
TABLA 30 <i>COMPARATIVA ENTRE MODELOS</i>	128
TABLA 31 <i>PARÁMETROS DE ENTRENAMIENTO TiDE</i>	131
TABLA 32 <i>TAMAÑO DE LOS MODELOS</i>	132
TABLA 33 <i>COMPARACIÓN ENTRE INSTANCIAS</i>	132
TABLA 34 <i>PREDICCIONES FUERA DE LA MUESTRA</i>	133
TABLA 35 <i>PREDICCIONES DEL MODELO TiDE SELECCIONADO</i>	133
TABLA 37 <i>COMPARACIÓN MAE ENTRE MODELOS ENTRENADOS</i>	141
TABLA 38 <i>EVALUACIÓN DE DESEMPEÑO PREDICTIVO</i>	143

CAPÍTULO 1:

1. Introducción

Definición del proyecto

El presente proyecto tiene como objetivo desarrollar y evaluar un esquema de pronóstico multi-horizonte para estimar el comportamiento del Índice de Tipo de Cambio Efectivo Real (ITCER) del Ecuador mediante técnicas de series temporales, aprendizaje automático y aprendizaje profundo. El estudio se centra en analizar si modelos avanzados, incluidas arquitecturas de aprendizaje profundo basadas en Transformers para series temporales, pueden mejorar el desempeño predictivo frente a modelos base tradicionales en horizontes de corto y mediano plazo.

La economía de Ecuador es dolarizada, donde el tipo de cambio nominal no puede utilizarse como mecanismo directo de ajuste macroeconómico. En este escenario, el cambio del ITCER depende principalmente de factores asociados a precios relativos, inflación, condiciones externas y variables macroeconómicas relevantes. Debido a estas influencias, resulta pertinente evaluar modelos capaces de integrar múltiples covariables y capturar relaciones temporales no lineales presentes en los datos económicos.

La relevancia del estudio radica en la importancia del ITCER como indicador de competitividad externa del país. Un sistema de pronóstico más preciso puede contribuir al análisis económico y al seguimiento de factores que afectan la competitividad internacional del Ecuador, proporcionando información útil para investigadores, analistas y tomadores de decisiones.

Para abordar este problema, la investigación utilizará una estrategia escalonada de modelado. En una primera etapa se implementarán modelos base de referencia, incluyendo un modelo ingenuo y un modelo estadístico con variables exógenas. Posteriormente, se

evaluarán modelos de aprendizaje automático y aprendizaje profundo enfocados en series temporales para comparar su desempeño frente a las líneas base bajo los mismos horizontes de predicción y métricas de evaluación.

Dentro de los modelos avanzados se considerarán arquitecturas con esquemas de trabajo sobre pronóstico multi-horizonte y permitan incorporar variables exógenas, como LightGBM, LSTM, Temporal Fusion Transformer (TFT) y TiDE. El modelo TFT se incluye por su capacidad para realizar predicciones multi-horizonte, integrar covariables y sus mecanismos internos de interpretabilidad. La comparación buscará identificar qué modelo ofrece mejor desempeño predictivo y utilidad interpretativa para el caso del ITCER ecuatoriano.

La investigación se enfocará en un esquema de pronóstico multi-horizonte simultáneo para horizontes de 1, 3 y 6 meses. El conjunto de variables exógenas se limitará exclusivamente a cinco indicadores macroeconómicos: IPC de Ecuador, IPC de Estados Unidos, USD NEER, precio del petróleo WTI y riesgo país EMBI, además de variables de calendario asociadas a la temporalidad mensual. Estas variables fueron indentificadas por su relación directa con precios relativos, condiciones externas y factores macroeconómicos que pueden influir en la evolución del ITCER.

El desarrollo metodológico comprenderá cuatro etapas principales. Primero, se recopilarán, depurarán y alinearán en una sola frecuencia temporal las series históricas. Luego, se construirá el conjunto de datos supervisado para el problema de pronóstico multi-horizonte. Posteriormente, se implementarán y evaluarán los modelos seleccionados mediante métricas de error como MAE, RMSE y sMAPE, con el fin de comparar su desempeño en los horizontes de 1, 3 y 6 meses. Finalmente, se analizará la interpretabilidad de los modelos que

permitan este tipo de evaluación e identificar la contribución de las variables y su relación con las predicciones.

Justificación e importancia del trabajo de investigación

El análisis del Índice de Tipo de Cambio Efectivo Real es importante para evaluar la competitividad externa de Ecuador, ya que, al ser una economía dolarizada, el país no puede ajustar su tipo de cambio nominal y depende del ITCER para medir su relación de precios con sus socios comerciales. Además, un pronóstico más preciso del ITCER permite anticipar cambios en la competitividad del país y apoyar la toma de decisiones en materia de política económica y planificación empresarial, especialmente en sectores vinculados al comercio exterior. En este sentido, el uso de modelos de aprendizaje profundo, incluidas arquitecturas basadas en Transformers para series temporales como el Temporal Fusion Transformer, permite modelar relaciones no lineales y dependencias temporales entre variables macroeconómicas, lo que puede lograr en una mejora en la precisión de los pronósticos.

El uso del TFT se justifica porque el pronóstico del ITCER involucra múltiples covariables macroeconómicas y no solo una serie univariada. El modelo permite integrar información histórica, variables exógenas y covariables futuras dentro de una arquitectura de predicción multi-horizonte, además de incorporar mecanismos de atención temporal e interpretabilidad.

La presente investigación contribuye al ámbito académico mediante la integración de técnicas de aprendizaje profundo y herramientas de interpretabilidad aplicadas al análisis de series temporales económicas. Además, el desarrollo del proyecto resulta factible gracias a la disponibilidad de información proveniente de fuentes oficiales, como el Banco Central del Ecuador, el Instituto Nacional de Estadística y Censos (INEC) y el Ministerio de Economía y Finanzas.

Alcance

La presente investigación se enfoca en el pronóstico mensual del Índice de Tipo de Cambio Efectivo Real del Ecuador mediante modelos de series temporales y aprendizaje profundo. El estudio abordará el problema desde un enfoque multi-horizonte, considerando predicciones a 1, 3 y 6 meses, con el propósito de evaluar el comportamiento futuro del ITCER en el corto y mediano plazo utilizando la información correspondiente al periodo 2018-2026. Esta selección se justifica por la disponibilidad de datos de todas las variables exógenas y porque el Banco Central del Ecuador adoptó el año 2018 como el año base de referencia actual (Base 2018 = 100) para el cálculo ponderado del índice frente a sus 16 socios comerciales.

El análisis utilizará como variable objetivo la serie histórica mensual del ITCER publicada por el Banco Central del Ecuador. Esta variable será complementada con un conjunto definido de variables macroeconómicas exógenas relacionadas con la dinámica de precios relativos, condiciones externas y entorno económico y financiero. Las variables consideradas en el estudio serán el IPC de Ecuador, el IPC de Estados Unidos, el USD NEER, el precio mensual del petróleo WTI y el riesgo país EMBI. Adicionalmente, se incorporarán variables de calendario, como mes, trimestre y tendencia temporal, para representar posibles patrones estacionales y efectos temporales presentes en la serie.

Las series utilizadas serán estructuradas a frecuencia mensual. Para el caso de las variables que se encuentren en una frecuencia distinta, se realizará su transformación a una granularidad mensual, manteniendo una base histórica homogénea para el entrenamiento, validación y evaluación de los modelos.

La estructura de los datos será multi-horizonte de tipo simultáneo. Es decir, los modelos utilizarán una ventana histórica de observaciones pasadas para generar predicciones

directas sobre los horizontes de 1, 3 y 6 meses. Esta estructura permitirá evaluar el desempeño predictivo de cada modelo en distintos horizontes temporales sin usar una estrategia recursiva de predicción.

El estudio plantea una metodología de modelado incremental. En una primera fase, se implementarán modelos base que servirán como referencia para la comparación de resultados. Estos modelos permitirán establecer un punto de partida sobre el desempeño esperado en el pronóstico del ITCER. Posteriormente, se evaluarán modelos de mayor complejidad, incluyendo enfoques de aprendizaje automático y aprendizaje profundo, con el fin de determinar si estos aportan mejoras frente a los métodos tradicionales.

Esta secuencia permitirá comparar el rendimiento de distintos tipos de modelos bajo las mismas condiciones de evaluación, considerando los horizontes de predicción de 1, 3 y 6 meses. La precisión del pronóstico se analizará de forma separada para cada horizonte, debido a que el error puede variar conforme aumenta la distancia temporal de la predicción.

Dentro de los modelos de aprendizaje profundo se incluirá el Temporal Fusion Transformer (TFT), debido a que permite trabajar con series históricas, covariables dinámicas y variables futuras conocidas. Además, este modelo incorpora mecanismos de selección de variables y atención temporal, los cuales permiten analizar la relevancia relativa de los predictores y de los rezagos utilizados en las estimaciones. El TFT será evaluado junto con otros modelos seleccionados. Debido a su complejidad se utilizará una implementación optimizada para series temporales con pocos datos.

Para la validación de los modelos, se emplearán métricas de error consolidadas en el análisis de series temporales, tales como MAE, RMSE y MAPE. Estos indicadores se calcularán de manera desagregada por horizonte, permitiendo contrastar la eficacia de las

arquitecturas tanto a nivel agregado como en función del incremento en la distancia temporal del pronóstico.

El estudio se limitará al desarrollo, entrenamiento, evaluación e interpretación de modelos predictivos en un entorno de investigación aplicada. No se contempla la implementación del modelo en un sistema productivo, la generación de predicciones en tiempo real ni la formulación de recomendaciones en política económica. Los resultados estarán orientados a evaluar la capacidad predictiva e interpretativa de los modelos seleccionados para el caso específico del ITCER en Ecuador.

Objetivos

1.1.1. Objetivo General

Desarrollar y evaluar un esquema de pronóstico multi-horizonte del Índice de Tipo de Cambio Efectivo Real (ITCER) del Ecuador mediante modelos de series temporales, aprendizaje automático y aprendizaje profundo, incorporando variables macroeconómicas exógenas y evaluación de interpretabilidad para comparar su desempeño predictivo en horizontes de 1, 3 y 6 meses.

1.1.2. Hipótesis

La incorporación de variables macroeconómicas exógenas en modelos de pronóstico multi-horizonte permitirá mejorar el desempeño predictivo del Índice de Tipo de Cambio Efectivo Real del Ecuador frente a modelos base univariados o tradicionales. Esta mejora se evaluará mediante métricas de error como MAE, RMSE y sMAPE en horizontes de 1, 3 y 6 meses. Además, se espera que los modelos con mecanismos de interpretabilidad permitan identificar la contribución relativa de las variables macroeconómicas utilizadas en las predicciones generadas.

1.1.3. Objetivos Específicos

- Recopilar, depurar y alinear temporalmente las series históricas mensuales del Índice de Tipo de Cambio Efectivo Real (ITCER) y de las variables macroeconómicas exógenas seleccionadas para la construcción del conjunto de datos del estudio.
- Construir un esquema supervisado de pronóstico multi-horizonte, definiendo ventanas históricas de entrada, horizontes de predicción de 1, 3 y 6 meses, particiones temporales de entrenamiento, validación y prueba, y métricas de evaluación del desempeño predictivo.
- Implementar modelos base de comparación, incluyendo modelos ingenuos y modelos estadísticos con variables exógenas, como referencia para evaluar el desempeño de los modelos más complejos.
- Implementar y entrenar modelos de aprendizaje automático y aprendizaje profundo, incorporando variables macroeconómicas exógenas y covariables de calendario para el pronóstico multi-horizonte del ITCER.
- Comparar el desempeño de los modelos seleccionados mediante métricas de error de pronóstico multi-horizonte, con el fin de evaluar si los enfoques de mayor complejidad aportan mejoras predictivas cuantificables frente a los modelos base.
- Comparar el desempeño del modelo Temporal Fusion Transformer frente a los modelos base seleccionados mediante métricas de error de pronóstico multi-horizonte, con el fin de evaluar si el modelo avanzado aporta mejoras predictivas cuantificables.
- Analizar la interpretabilidad del modelo Temporal Fusion Transformer para identificar las variables exógenas y los segmentos temporales con mayor

contribución en las predicciones del ITCER para los distintos horizontes analizados.

CAPÍTULO 2:

2. Revisión de Literatura

2.1 Estado del Arte

El Índice de Tipo de Cambio Efectivo Real (ITCER) es un indicador utilizado para evaluar la competitividad externa de una economía, al relacionar el nivel de precios internos con el de sus principales socios comerciales y considerar el efecto del tipo de cambio. En términos generales, un aumento del ITCER se interpreta como una mejora de la competitividad, mientras que una disminución refleja un deterioro relativo. En economías con moneda propia, la evolución del indicador está determinada tanto por el tipo de cambio nominal como por las diferencias de inflación respecto de los socios comerciales, por lo que las variaciones cambiarias pueden contribuir al proceso de ajuste de la competitividad.

En Ecuador, el análisis del ITCER presenta características particulares debido al régimen de dolarización vigente. Al utilizar el dólar estadounidense como moneda de curso legal, el país no dispone de una política cambiaria autónoma ni puede modificar directamente el tipo de cambio nominal para compensar diferencias de inflación. En consecuencia, la evolución del ITCER depende principalmente de la relación entre los precios internos y los precios de los socios comerciales. En este contexto, el Banco Central del Ecuador calcula el índice en su versión multilateral (ITCER) y mediante índices bilaterales (ITCERB), utilizando, entre otros elementos, índices de precios al consumidor. Esta característica hace que la competitividad externa de Ecuador dependa en mayor medida de factores como la evolución de los precios, la productividad y los costos internos.

La ausencia de mecanismos de ajuste cambiario convierte al seguimiento de la competitividad externa en un aspecto relevante para el análisis económico del país. En este sentido, el pronóstico del ITCER puede contribuir a anticipar cambios en su trayectoria y

proporcionar información útil para la evaluación de las condiciones de competitividad de la economía ecuatoriana.

En el ámbito del pronóstico de series temporales, los modelos autorregresivos integrados de media móvil (ARIMA) constituyen uno de los enfoques estadísticos más utilizados. Un modelo autorregresivo establece que el valor de una variable en un determinado período puede explicarse, en parte, a partir de sus propios valores observados en períodos anteriores, incorporando un término de error que representa la información no explicada por el modelo (De Arce et al., 2023). Bajo determinados supuestos, los procesos estocásticos pueden representarse mediante una combinación de sus valores pasados y un término de innovación aleatoria, lo que permite modelar la dependencia temporal presente en una serie.

Una extensión de este enfoque corresponde a los modelos SARIMA, diseñados para series temporales que presentan patrones estacionales. Estos modelos incorporan componentes autorregresivos, de diferenciación y de medias móviles tanto para la parte no estacional como para la estacional de la serie. De acuerdo con Hyndman y Athanasopoulos (2021), esta estructura permite representar simultáneamente tendencias, dependencia temporal y patrones que se repiten a intervalos regulares, como ocurre en series con frecuencia mensual, trimestral o anual.

Mientras que los modelos ARIMA se enfocan principalmente en una única variable, los modelos de vectores autorregresivos (VAR) permiten analizar de manera conjunta varias series temporales. Su principal característica consiste en representar cada variable como una función de sus propios valores rezagados y de los valores pasados de las demás variables que conforman el sistema. De esta manera, un modelo VAR con n variables se estructura mediante un sistema de n ecuaciones, permitiendo estudiar las interacciones dinámicas entre

diferentes variables económicas. Esta característica resulta relevante en contextos macroeconómicos, donde la evolución de una variable puede estar relacionada con el comportamiento pasado de otras variables del sistema.

Otro enfoque estadístico ampliamente utilizado en el pronóstico de series temporales es el método de Holt-Winters. Esta técnica de suavizamiento exponencial permite modelar series que presentan tendencia y estacionalidad, asignando mayor peso a las observaciones recientes. Hyndman y Athanasopoulos (2021) señalan que el método incorpora componentes asociados al nivel, la tendencia y la estacionalidad de la serie, por lo que constituye una alternativa adecuada para generar pronósticos cuando existen patrones temporales relativamente estables.

A pesar de la amplia utilización de estos métodos estadísticos, la literatura reciente ha impulsado el desarrollo de enfoques capaces de representar relaciones no lineales y estructuras más complejas. Los modelos tradicionales, como ARIMA y VAR, dependen de determinados supuestos sobre la estructura de la serie y pueden presentar limitaciones cuando existen relaciones no lineales, cambios estructurales o interacciones complejas entre múltiples variables. Estas características son especialmente relevantes en los mercados cambiarios, donde pueden producirse modificaciones en los patrones de comportamiento como consecuencia de shocks económicos, cambios en los precios internacionales o modificaciones en las condiciones financieras (Pirayesh Neghab et al., 2025).

En respuesta a estas limitaciones, los métodos de aprendizaje automático han adquirido mayor relevancia en el pronóstico de variables económicas y financieras. Algoritmos como Random Forest (RF) permiten modelar relaciones no lineales y considerar simultáneamente un conjunto amplio de variables explicativas. Asimismo, métodos de potenciación de gradiente, como CatBoost, han mostrado resultados favorables en problemas

de predicción de variables macrofinancieras, debido a su capacidad para representar relaciones complejas entre las variables de entrada (Pirayesh Neghab et al., 2025).

Un desarrollo relacionado con estos modelos corresponde a la Inteligencia Artificial Explicable (Explainable Artificial Intelligence, XAI), cuyo objetivo es proporcionar mecanismos que permitan interpretar los resultados de modelos de aprendizaje automático. En este ámbito, las técnicas basadas en SHAP (SHapley Additive exPlanations) permiten estimar la contribución de cada variable a una determinada predicción. Pirayesh Neghab et al. (2025) emplean este tipo de herramientas para analizar los determinantes de las predicciones cambiarias, mostrando que variables relacionadas con los términos de intercambio, entre ellas los precios de las materias primas, pueden ejercer una influencia relevante sobre el comportamiento proyectado del tipo de cambio.

De acuerdo con Rob J Hyndman y George Athanasopoulos (2021), el método Holt-Winters extiende el suavizamiento exponencial simple incorporando componentes de tendencia y estacionalidad, lo que lo convierte en una herramienta adecuada para pronósticos de corto y mediano plazo en series temporales con patrones periódicos.

Tradicionalmente, los modelos se han sustentado en enfoques estadísticos y econométricos clásicos, tales como los modelos autorregresivos integrados de media móvil (ARIMA) y de vectores autorregresivos (VAR). Sin embargo, existe un consenso, en la literatura en que estas técnicas clásicas están limitadas por supuestos restrictivos de linealidad y estacionariedad que les impiden capturar con éxito la no linealidad, el ruido estocástico y las rupturas estructurales de los mercados de divisas (Pirayesh Neghab et al., 2025).

Frente a las limitaciones de los modelos paramétricos tradicionales, la literatura reciente ha incorporado algoritmos de aprendizaje supervisado capaces de modelar relaciones no lineales entre variables macroeconómicas sin requerir una especificación funcional previa.

Dentro de este grupo se encuentran modelos basados en árboles de decisión y métodos de ensamble, como Random Forest, XGBoost, LightGBM y CatBoost, los cuales han sido utilizados en problemas de pronóstico económico y financiero por su capacidad para capturar interacciones entre variables y manejar conjuntos de predictores de distinta naturaleza.

En el caso del pronóstico cambiario, Pirayesh Neghab et al. (2025) proponen un enfoque de aprendizaje automático interpretable para analizar el tipo de cambio entre el dólar canadiense y el dólar estadounidense a partir de fundamentos macroeconómicos. Su estudio muestra que los modelos de aprendizaje automático pueden ser útiles no solo para mejorar el pronóstico, sino también para examinar la contribución de variables económicas específicas. En particular, los autores identifican el precio del petróleo como una variable relevante para explicar la dinámica cambiaria canadiense, debido a la importancia de este commodity en su economía.

En esta línea, los modelos basados en árboles resultan pertinentes para el presente estudio, ya que permiten incorporar variables macroeconómicas exógenas y evaluar relaciones no lineales con el ITCER. Entre estos modelos, LightGBM es una alternativa relevante porque implementa árboles de decisión potenciados por gradiente con mejoras orientadas a reducir el tiempo de entrenamiento y manejar conjuntos de variables de manera eficiente. Ke et al. (2017) señalan que LightGBM introduce técnicas como Gradient-based One-Side Sampling y Exclusive Feature Bundling, las cuales permiten acelerar el entrenamiento sin perder capacidad predictiva de forma significativa.

Un aspecto importante de estos modelos es que pueden complementarse con herramientas de Inteligencia Artificial Explicable, como SHAP. Este enfoque permite estimar la contribución de cada variable a las predicciones del modelo y facilita una lectura más clara de los resultados. En el contexto de esta investigación, el uso de interpretabilidad resulta útil

para analizar si variables como el IPC de Ecuador, el IPC de Estados Unidos, el USD NEER, el WTI y el EMBI aportan información relevante al pronóstico del ITCER en distintos horizontes temporales.

Por tanto, la inclusión de modelos como LightGBM en la comparación permite evaluar si un modelo supervisado de menor complejidad que una red profunda puede obtener un desempeño competitivo en una serie mensual con pocas observaciones. Además, su compatibilidad con métodos de interpretación de variables permite comparar no solo el error de pronóstico, sino también la coherencia económica de los predictores utilizados.

Paralelamente, investigaciones recientes de Ding et al. (2024) han ampliado este espectro mediante la fusión de datos estructurados cuantitativos y datos textuales no estructurados (noticias financieras y análisis de políticas) procesados a través de Grandes Modelos de Lenguaje (LLM), demostrando que la incorporación de sentimientos de mercado reduce considerablemente el error absoluto medio (MAE) y el error cuadrático medio (RMSE) en el pronóstico de divisas globales frente a modelos univariados convencionales.

En el plano de los modelos de aprendizaje profundo (Deep Learning), las redes neuronales recurrentes, específicamente las redes de Memoria a Corto y Largo Plazo (LSTM) y las Unidades Recurrentes Gated (GRU), representaron el primer gran avance para procesar la naturaleza secuencial y temporal de las series macroeconómicas y financieras, gracias a sus compuertas internas diseñadas para regular la persistencia y resolver el desvanecimiento del gradiente. Sin embargo, la investigación aplicada se encuentra dominada actualmente por las arquitecturas basadas en el mecanismo de autoatención como los Transformers (Zeng et al., 2023).

En los últimos años, las arquitecturas basadas en Transformers han emergido como una alternativa sólida a los modelos tradicionales de series temporales, debido a su capacidad

para capturar dependencias de largo plazo y relaciones no lineales complejas. Originalmente propuestos en el contexto del procesamiento de lenguaje natural por Ashish Vaswani et al. (2017), estos modelos han sido adaptados al análisis temporal, dando lugar a los denominados Transformers temporales. A diferencia de modelos como LSTM o GRU, los Transformers utilizan mecanismos de autoatención que permiten ponderar dinámicamente la relevancia de diferentes puntos en el tiempo, lo que resulta especialmente útil en problemas de pronóstico multi-horizonte, donde es necesario predecir múltiples pasos hacia adelante de manera simultánea.

Los Transformers aplicados a series temporales pueden clasificarse en varias categorías según su diseño y propósito. En primer lugar, se encuentran los modelos basados en encoder-decoder, que siguen la arquitectura clásica y son ampliamente utilizados para predicciones multi-horizonte, como el Temporal Fusion Transformer (TFT) propuesto por Bryan Lim et al. (2021). En segundo lugar, existen modelos eficientes o de atención reducida, como Informer y Autoformer, que buscan disminuir el costo computacional del mecanismo de atención en series largas mediante técnicas de sparsity o descomposición. En tercer lugar, se encuentran los modelos híbridos, que combinan Transformers con redes recurrentes o convolucionales para mejorar la extracción de patrones locales y globales. Finalmente, se han desarrollado modelos especializados en interpretabilidad, que incorporan mecanismos explícitos para explicar la contribución de variables y tiempos en la predicción.

El uso de Transformers temporales es de mucha utilidad en contextos donde las series presentan alta complejidad, múltiples variables y relaciones dinámicas en el tiempo, como son el caso de indicadores macroeconómicos. En el pronóstico del ITCER, por ejemplo, es fundamental captar la interacción entre variables como USD NEER, precio del petróleo, balanza comercial, reservas internacionales y riesgo país. Los mecanismos de atención permiten identificar qué períodos históricos y qué variables tienen mayor influencia en cada

horizonte de predicción, lo cual supera a los modelos tradicionales que dependen de supuestos de linealidad o estacionariedad.

Un aspecto clave en la adopción de Transformers en aplicaciones económicas es la interpretabilidad. Aunque estos modelos son altamente precisos, suelen considerarse “cajas negras”. Para abordar este problema, se han desarrollado enfoques como el TFT, que integra capas de atención interpretables y mecanismos de selección de variables, permitiendo analizar la importancia de cada entrada en diferentes horizontes temporales. Esta característica es relevante en estudios económicos, donde no solo se requiere o importa la predicción, sino también la explicación del comportamiento de las variables.

Por lo anterior, varios modelos de Transformers temporales pueden ser considerados para la presente investigación sobre el pronóstico multi-horizonte del ITCER utilizando datos del Banco Central del Ecuador. Entre los más relevantes se encuentran: (i) el Temporal Fusion Transformer (TFT), por su enfoque en interpretabilidad y manejo de variables estáticas y dinámicas; (ii) Informer, por su eficiencia en series largas; (iii) Autoformer, por su capacidad de descomposición de tendencias y estacionalidad; y (iv) modelos más recientes que integran atención interpretable y aprendizaje multitarea. La elección del modelo dependerá del equilibrio entre precisión, interpretabilidad y disponibilidad de datos, siendo especialmente recomendable el uso de enfoques interpretables en contextos macroeconómicos.

En este contexto, el Temporal Fusion Transformer (TFT), propuesto por Lim et al. (2021), constituye una arquitectura relevante para el pronóstico multi-horizonte de series temporales. Este modelo combina mecanismos de atención, selección de variables y componentes recurrentes, lo que le permite trabajar con relaciones temporales de corto y largo plazo. Además, ofrece herramientas de interpretabilidad que ayudan a revisar la

importancia relativa de las variables y ciertos patrones temporales asociados a las predicciones. Por estas razones, el TFT resulta pertinente como modelo de referencia para el análisis del ITCER, ya que este indicador puede estar influido por distintas variables macroeconómicas y por efectos temporales no lineales.

De manera adicional, investigaciones recientes han desarrollado variantes de Transformers temporales orientadas a mejorar la interpretación de series multivariadas. Por ejemplo, Liu et al. (2024) proponen el Interpretable Feature-Temporal Transformer (IFTT) para el pronóstico de energía eólica, incorporando mecanismos de atención sobre variables y sobre la dimensión temporal. Aunque este estudio se aplica al sector energético, resulta útil como referencia metodológica porque muestra cómo los modelos de pronóstico pueden combinar precisión e interpretabilidad.

También, Chen et al. (2024) presentan Pathformer, un modelo basado en Transformers multi-escala con rutas adaptativas. A diferencia de enfoques que procesan la serie desde una única escala temporal, Pathformer divide la secuencia en segmentos de diferentes tamaños, lo que permite representar patrones de corto y largo plazo. Además, utiliza mecanismos de atención para capturar relaciones locales y globales entre los distintos segmentos de la serie.

El aporte principal de Pathformer se encuentra en dos componentes: el bloque Transformer multi-escala y el mecanismo de rutas adaptativas. El primero permite trabajar con distintas resoluciones temporales, mientras que el segundo selecciona y combina la información más relevante según las características de los datos. Este enfoque es importante para el análisis de series macroeconómicas, ya que indicadores como el ITCER pueden presentar cambios de corto plazo, tendencias graduales y efectos asociados a choques externos. No obstante, en esta investigación Pathformer se considera como una referencia del

estado del arte, y no como evidencia directa sobre el desempeño del modelo en el pronóstico del ITCER.

Por otro lado, el uso de transformers temporales también presenta limitaciones que deben ser consideradas. En primer lugar, este tipo de modelos suele requerir una cantidad grande de datos históricos para lograr un aprendizaje estable, por lo que su precisión y comportamiento puede verse afectado cuando la serie disponible no es lo suficientemente larga o presenta cambios metodológicos a lo largo del tiempo. De igual manera, su implementación demanda un mayor esfuerzo computacional en comparación con modelos más simples, ya que requiere definir y ajustar múltiples hiperparámetros, estructuras de entrada y configuraciones de entrenamiento, lo cual puede incrementar el tiempo de desarrollo del modelo.

Asimismo, la aplicación de un transformer temporal no garantiza mejores resultados que otros enfoques de menor complejidad. En problemas de pronóstico de series de tiempo, especialmente cuando la dinámica de la serie no es compleja, modelos más tradicionales pueden alcanzar un alto desempeño con una validación más sencilla. También, al incorporar variables exógenas es necesario asegurar una adecuada alineación temporal, distinguir correctamente entre variables observadas históricamente y variables conocidas a futuro, y evitar filtraciones de información en la etapa de entrenamiento y evaluación.

El Temporal Fusion Transformer (TFT), propuesto por Bryan Lim y colaboradores, corresponde a una arquitectura encoder-decoder basada en Transformers temporales, diseñada específicamente para problemas de pronóstico multi-horizonte con múltiples variables. A diferencia de modelos tradicionales de series de tiempo, el TFT permite integrar simultáneamente información histórica de la variable objetivo, variables exógenas dinámicas y covariables conocidas a futuro, como variables de calendario. Su estructura combina

mecanismos de atención temporal, redes recurrentes y módulos de selección de variables, permitiendo capturar dependencias temporales complejas y relaciones no lineales presentes en series macroeconómicas. Estas características lo convierten en un enfoque adecuado para el pronóstico del ITCER, donde intervienen múltiples factores económicos interrelacionados y distintos horizontes temporales de predicción.

Uno de los principales aportes del TFT es su capacidad de interpretabilidad, aspecto especialmente relevante en aplicaciones económicas y financieras. El modelo incorpora mecanismos explícitos de atención y selección de variables que permiten identificar qué variables y qué segmentos temporales tienen mayor influencia en las predicciones generadas para cada horizonte. Esto representa una ventaja frente a otros modelos de aprendizaje profundo considerados “cajas negras”, ya que no solo se busca obtener precisión predictiva, sino también comprender los factores que explican el comportamiento del ITCER. No obstante, el uso del TFT también presenta limitaciones asociadas a su complejidad computacional, necesidad de ajuste de hiperparámetros y dependencia de una adecuada preparación de datos, por lo que su implementación debe evaluarse considerando la disponibilidad de información histórica y los recursos computacionales del proyecto.

Finalmente, tenemos El Time-series Dense Encoder (TiDE) que fue propuesto por Das et al. (2023) como una arquitectura de aprendizaje profundo para pronósticos de múltiples pasos. El modelo surgió en un contexto en el que métodos lineales, como DLinear, habían obtenido resultados competitivos frente a modelos Transformer en tareas de pronóstico de largo horizonte. TiDE buscó conservar la eficiencia de dichos métodos, pero incorporando capacidad para representar relaciones no lineales y utilizar variables exógenas. El trabajo fue publicado en Transactions on Machine Learning Research.

A diferencia de las redes recurrentes y de los modelos basados en atención, TiDE está formado principalmente por perceptrones multicapa o MLP. No utiliza mecanismos de recurrencia, convoluciones ni autoatención. Por esta razón, su costo computacional crece de forma aproximadamente lineal con la longitud de la ventana histórica y del horizonte de pronóstico, mientras que la atención convencional puede presentar un crecimiento cuadrático respecto de la longitud de la secuencia (Das et al., 2023)

TiDE sigue una arquitectura de codificador y decodificador. Para cada serie, recibe tres tipos de información: los valores históricos de la variable objetivo, las covariables dinámicas y, cuando están disponibles, los atributos estáticos. La salida corresponde a todos los valores del horizonte de pronóstico:

$$\hat{\mathbf{y}}_{t+1:t+H} = [\hat{\mathbf{y}}_{\{t+1\}}, \hat{\mathbf{y}}_{\{t+2\}}, \dots, \hat{\mathbf{y}}_{\{t+H\}}]$$

Donde H representa el número de periodos futuros que se desea pronosticar y $\hat{\mathbf{y}}_{t+1:t+H}$ corresponde al vector de predicciones para dicho horizonte. Esta formulación permite producir el horizonte completo en una sola ejecución, sin utilizar cada predicción como entrada del periodo siguiente. Por tanto, TiDE aplica una estrategia directa de salida múltiple y evita la propagación de errores asociada con los métodos recursivos (Das et al., 2023).

En conjunto, la evolución de la literatura muestra una transición desde modelos estadísticos tradicionales hacia métodos de aprendizaje automático y arquitecturas de aprendizaje profundo capaces de representar relaciones no lineales y dependencias temporales complejas. Para el caso específico del ITCER ecuatoriano, esta evolución resulta relevante debido a la interacción de factores macroeconómicos, externos y financieros que pueden afectar su comportamiento. En este contexto, la comparación de modelos estadísticos, métodos de aprendizaje automático y arquitecturas de aprendizaje profundo como TFT, Pathformer y TiDE

permite evaluar distintas alternativas para el pronóstico multi-horizonte del indicador, considerando no solo la precisión de las predicciones, sino también la complejidad, interpretabilidad y disponibilidad de información.

2.2 Marco Teórico

2.2.1 Procesos estocásticos

Un proceso estocástico es un conjunto de variables aleatorias ordenadas en el tiempo, que permite representar fenómenos cuyo comportamiento futuro está sujeto a incertidumbre. Formalmente, un proceso estocástico puede expresarse como:

$$\{X_t\}_{t \in T}$$

Cada variable X_t representa los posibles valores que puede tomar un fenómeno en un periodo t , mientras que T corresponde al conjunto de instantes o períodos de tiempo considerados. Este concepto es fundamental para el estudio de series temporales, ya que permite modelar fenómenos en los que las observaciones futuras no pueden determinarse con certeza a partir de los valores presentes o pasados. En el ámbito económico, por ejemplo, el Índice de Tipo de Cambio Efectivo Real (ITCER) puede considerarse como una realización de un proceso estocástico, debido a que su comportamiento varía a lo largo del tiempo y está condicionado por diversos factores económicos.

2.2.2 Procesos estocásticos estacionarios

La estacionariedad constituye una propiedad fundamental en el análisis de series temporales. Un proceso se considera estacionario cuando sus principales propiedades estadísticas permanecen constantes a lo largo del tiempo. En particular, bajo el concepto de estacionariedad débil o de segundo orden, se requiere que la media y la varianza sean constantes y que la covarianza entre dos observaciones dependa únicamente de la distancia temporal que las separa. Estas condiciones pueden expresarse como:

$$E(X_t) = \mu, \quad Var(X_t) = \sigma^2 < \infty, \quad Cov(X_t, X_{\{t-k\}}) = \gamma_k$$

Donde X_t representa el valor del proceso en el período t , μ es una media constante, σ^2 es una varianza constante y finita, y γ_k representa la autocovarianza asociada al rezago k .

En consecuencia, la dependencia estadística entre dos observaciones no debe estar determinada por el momento específico en que estas ocurren, sino únicamente por la distancia temporal que existe entre ellas. La estacionariedad es relevante para el modelado y pronóstico de series temporales, debido a que facilita la identificación de relaciones temporales estables y permite aplicar diversos modelos estadísticos bajo supuestos definidos.

2.2.3 Series Temporales

Una serie temporal es un conjunto de observaciones de una variable registradas de forma secuencial a lo largo del tiempo. El orden cronológico constituye una característica esencial, debido a que las observaciones pueden presentar relaciones de dependencia con valores registrados en períodos anteriores. De manera general, una serie temporal puede representarse como:

$$Y = y_1, y_2, y_3, \dots, y_T$$

Donde y_t representa el valor observado de la variable en el período t , mientras que T corresponde al número total de observaciones disponibles.

Las series temporales tienen una amplia aplicación en el análisis económico, particularmente para estudiar variables como la inflación, el producto interno bruto, las tasas de interés, los precios de materias primas, los índices de precios y los tipos de cambio. Estas variables pueden presentar tendencias, patrones estacionales, ciclos y relaciones de dependencia entre observaciones sucesivas, además de verse afectadas por cambios en las condiciones económicas nacionales e internacionales.

En la presente investigación, la variable objetivo corresponde al Índice de Tipo de Cambio Efectivo Real (ITCER) de Ecuador. La serie se encuentra registrada con frecuencia mensual, por lo que cada observación representa el valor del índice correspondiente a un mes determinado. El período de estudio comprende los años 2018 a 2025, de acuerdo con la disponibilidad conjunta de la variable objetivo y de las variables macroeconómicas consideradas en el análisis.

Función de autocovarianza y autocorrelación

La autocovarianza permite medir la relación lineal entre una variable y su propio valor en un período anterior. Para un proceso estacionario, la autocovarianza en el rezago (k) se define como:

$$\gamma(k) = Cov(X_t, X_{t-k})$$

Por su parte, la función de autocorrelación expresa la relación existente entre las observaciones de una serie en diferentes períodos, estandarizando la autocovarianza respecto de la varianza:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Donde $\gamma(0)$ corresponde a la varianza del proceso. La autocorrelación toma valores en el intervalo $([-1,1])$ y permite identificar el grado de dependencia lineal entre una observación y sus valores rezagados.

Estas funciones son relevantes para determinar la estructura de dependencia temporal de una serie y constituyen herramientas fundamentales para la identificación y especificación de modelos de series temporales.

2.2.3.1 Proceso de ruido blanco

El ruido blanco representa un proceso aleatorio cuyos valores no presentan dependencia lineal entre diferentes períodos. En su forma habitual, se caracteriza por una media igual a cero, una varianza constante y una autocovarianza nula para cualquier rezago distinto de cero:

$$E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma^2$$

$$Cov(\varepsilon_t, \varepsilon_{t-k}) = 0, \quad k \neq 0$$

En el contexto de los modelos de series temporales, el término de ruido blanco representa las perturbaciones o innovaciones que no pueden ser explicadas por la estructura sistemática del modelo. Por ello, un modelo adecuadamente especificado debería dejar residuos que presenten características compatibles con un proceso de ruido blanco.

2.2.3.2 Componentes de una serie temporal

El comportamiento de una serie temporal puede analizarse a partir de diferentes componentes que permiten identificar sus principales patrones de evolución. Los componentes tradicionalmente considerados son la tendencia, la estacionalidad, el ciclo y el componente irregular o aleatorio.

Bajo una representación aditiva, la serie puede expresarse como:

$$y_t = T_t + S_t + C_t + \varepsilon_t$$

Donde:

- T_t : Representa la tendencia;
- S_t : Corresponde al componente estacional;
- C_t : Representa el comportamiento cíclico;

- ε_t : Recoge las variaciones irregulares o aleatorias.

La tendencia representa el comportamiento general de la serie durante un período prolongado. Puede presentar una trayectoria creciente, decreciente o relativamente estable y estar asociada con cambios estructurales o variaciones sostenidas en las condiciones económicas.

La estacionalidad corresponde a patrones que se repiten a intervalos regulares. En una serie mensual, por ejemplo, una periodicidad anual implica un ciclo de doce meses. De manera simplificada, la relación estacional puede expresarse como:

$$y_t \approx y_{t-s}$$

donde s representa la longitud del ciclo estacional. Para datos mensuales con un patrón anual, $s = 12$.

El componente cíclico representa fluctuaciones que se desarrollan durante períodos relativamente prolongados y cuya duración no necesariamente es fija. En el ámbito económico, estos movimientos pueden estar relacionados con fases de expansión, desaceleración y recuperación de la actividad económica.

Finalmente, el componente irregular recoge las variaciones que no pueden atribuirse directamente a la tendencia, la estacionalidad o el ciclo. Puede estar asociado con shocks externos, acontecimientos económicos o políticos inesperados, cambios en las condiciones del mercado o errores de medición.

2.2.3.3 Dependencia temporal

Una característica fundamental de las series temporales es la existencia de dependencia entre observaciones. El valor de una variable en un período determinado puede estar relacionado con sus valores registrados en períodos anteriores. Esta dependencia

temporal constituye uno de los fundamentos de los modelos autorregresivos y también es aprovechada por diversos métodos de aprendizaje automático.

De forma general, el valor de una serie puede representarse como una función de sus valores rezagados:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t$$

donde:

- y_t : Es el valor actual.
- $y_{t-1}, y_{t-2}, \dots, y_{t-p}$: Son valores anteriores de la serie.
- p : Representa el número de rezagos utilizados.
- $f(\cdot)$: es la relación estimada por el modelo.
- ε_t : Representa el error.

Cuando la función $f(\cdot)$ es lineal, puede utilizarse un modelo estadístico. Cuando las relaciones son más complejas o no lineales, se pueden aplicar métodos como LightGBM o modelos de aprendizaje profundo.

En esta investigación, los valores rezagados del ITCER constituyen una fuente de información sobre el comportamiento histórico del indicador. Adicionalmente, se incorporan variables macroeconómicas como el índice de precios al consumidor (IPC) de Ecuador, el IPC de Estados Unidos, el USD NEER, el precio del petróleo WTI y el riesgo país EMBI. También se consideran variables de calendario, como el mes, el trimestre y una variable de tendencia temporal. En términos generales, la relación entre la variable objetivo y las variables explicativas puede expresarse como:

$$y_t = f(y_{t-1}, \dots, y_{t-p}, x_{1,t}, x_{2,t}, \dots, x_{k,t}) + \varepsilon_t$$

Donde $x_{1,t}, x_{2,t}, x_{k,t}$ representan las variables explicativas disponibles en el período t .

2.2.4 Pronóstico de series temporales

El pronóstico de series temporales consiste en estimar valores futuros de una variable utilizando la información disponible hasta un determinado momento. Dependiendo de la cantidad y naturaleza de la información empleada, los modelos pueden clasificarse, entre otros criterios, en enfoques univariados y multivariados.

2.2.4.1 Series univariadas y multivariadas

En un enfoque univariado, el modelo utiliza exclusivamente la información histórica de la variable objetivo para generar las predicciones. De manera general:

$$\hat{y}_{\{t+h\}} = f(y_t, y_{\{t-1\}}, \dots, y_{\{t-p\}})$$

Donde (h) representa el horizonte de pronóstico y (p) el número de rezagos utilizados.

En un enfoque multivariado, además de la historia de la variable objetivo, se incorporan variables explicativas que pueden proporcionar información adicional sobre su comportamiento futuro:

$$\hat{y}_{\{t+h\}} = f(y_t, y_{\{t-1\}}, \dots, y_{\{t-p\}}, X_t)$$

Donde X_t representa el conjunto de variables explicativas.

En la presente investigación, el modelo ingenuo de persistencia constituye una referencia univariada, debido a que utiliza únicamente el último valor observado del ITCER. En contraste, modelos como SARIMAX, LightGBM y Temporal Fusion Transformer (TFT) pueden incorporar información adicional proveniente de variables macroeconómicas y de calendario.

2.2.4.2 Ventanas temporales

Los modelos de pronóstico requieren definir la cantidad de información histórica que será utilizada para generar una predicción. Esta cantidad se denomina ventana temporal, ventana histórica o ventana de observación.

Si se utiliza una ventana de longitud (L), la información histórica de la variable objetivo disponible en el período (t) puede representarse como:

$$\mathbf{X}_t = [y_{\{t-L+1\}}, y_{\{t-L+2\}}, \dots, y_t]$$

Cuando se incorporan variables explicativas, cada instante de la ventana contiene un vector de características:

$$\mathbf{X}_t = \begin{bmatrix} y_{t-L+1} & x_{1,t-L+1} & \cdots & x_{k,t-L+1} \\ y_{t-L+2} & x_{1,t-L+2} & \cdots & x_{k,t-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ y_t & x_{1,t} & \cdots & x_{k,t} \end{bmatrix}$$

En la presente investigación se plantea utilizar una ventana histórica de aproximadamente doce meses para la configuración ligera del modelo TFT. Esta elección considera la frecuencia mensual de la serie y el tamaño relativamente reducido de la muestra. Una ventana excesivamente extensa disminuiría el número de observaciones disponibles para construir muestras de entrenamiento y podría dificultar el aprendizaje del modelo.

2.2.4.3 Pronóstico multi-horizonte

El pronóstico multi-horizonte consiste en estimar varios valores futuros de una serie a partir de la información disponible hasta un período determinado. Si y_t representa la última observación conocida, una predicción realizada h períodos hacia adelante puede expresarse como:

$$\hat{y}_{\{t+h|t\}}$$

Donde: t corresponde al último período observado, h representa el horizonte de predicción y $\hat{y}_{\{t+h|t\}}$ es la estimación del valor futuro usando la información disponible hasta t .

Cuando $h = 1$, el modelo realiza una predicción para el siguiente período. Cuando se estiman varios valores futuros, el problema se denomina pronóstico multi-horizonte.

En esta investigación se consideran los horizontes de uno, tres y seis meses:

$$H = 1,3,6$$

Por tanto, la salida esperada del modelo se expresa como:

$$\hat{y}_t = [\hat{y}_{\{t+1|t\}}, \hat{y}_{\{t+3|t\}}, \hat{y}_{\{t+6|t\}}]$$

Estos horizontes permiten analizar el comportamiento inmediato del ITCER y observar cómo cambia la capacidad predictiva a medida que aumenta la distancia temporal.

2.2.4.4 Pronóstico de un paso

El pronóstico de un paso estima únicamente el valor inmediatamente posterior al último dato observado:

$$\hat{y}_{\{t+1\}} = f(X_t)$$

Este enfoque suele ser más sencillo porque la predicción se realiza para un período cercano. Sin embargo, no responde de manera directa a problemas en los que se requieren estimaciones para distintos momentos futuros.

2.2.4.5 Pronóstico de múltiples pasos

En el pronóstico de múltiples pasos, el modelo genera estimaciones para varios períodos futuros. Esta tarea puede realizarse mediante diferentes estrategias. Las más utilizadas son la estrategia recursiva, la estrategia directa y la estrategia de salida múltiple.

2.2.4.6 Estrategia recursiva

La estrategia recursiva utiliza un modelo entrenado para predecir un paso hacia adelante. Una vez obtenida la primera predicción, esta se incorpora como entrada para estimar el período siguiente.

La primera predicción se calcula como:

$$\hat{y}_{\{t+1\}} = f(y_t, y_{\{t-1\}}, \dots)$$

La segunda se obtiene utilizando la predicción anterior:

$$\hat{y}_{\{t+2\}} = f(\hat{y}_{\{t+1\}}, y_t, y_{\{t-1\}}, \dots)$$

De manera general:

$$\hat{y}_{\{t+h\}} = f(\hat{y}_{\{t+h-1\}}, \dots)$$

La ventaja de esta estrategia es que requiere un solo modelo, y su desventaja es la acumulación de errores. Cuando una predicción inicial contiene un error en la estimación, dicho error puede trasladarse a los pasos posteriores.

2.2.4.7 Estrategia directa

La estrategia directa emplea un modelo diferente para cada horizonte. Cada modelo aprende a estimar un valor futuro específico:

$$\hat{y}_{\{t+1\}} = f_1(\mathbf{X}_t)$$

$$\hat{y}_{\{t+3\}} = f_3(\mathbf{X}_t)$$

$$\hat{y}_{\{t+6\}} = f_6(\mathbf{X}_t)$$

La ventaja principal es que los errores no se transmiten de un horizonte a otro. Cada modelo se ajusta según el comportamiento propio del horizonte que debe estimar. Su desventaja es que requiere entrenar varios modelos. Por lo tanto, cada modelo utiliza como entrada los rezagos del ITCER, las variables macroeconómicas y las variables de calendario.

2.2.4.8 Estrategia de salida múltiple

La estrategia de salida múltiple utiliza un solo modelo para generar varias predicciones al mismo tiempo:

$$\hat{y}_t = f(\mathbf{X}_t)$$

con:

$$\hat{y}_t = [\hat{y}_{\{t+1\}}, \hat{y}_{\{t+3\}}, \hat{y}_{\{t+6\}}]$$

Este enfoque permite que el modelo aprenda relaciones compartidas entre los diferentes horizontes. Por ejemplo, las variables que ayudan a explicar el comportamiento del ITCER a un mes también pueden contener información útil para los horizontes de tres y seis meses.

El Temporal Fusion Transformer pertenece a este tipo de modelos. Su arquitectura está diseñada para integrar información histórica, covariables dinámicas y variables conocidas a futuro dentro de un esquema de pronóstico multi-horizonte.

2.2.4.9 Construcción de muestras multi-horizonte

Para entrenar los modelos es necesario transformar la serie temporal en un conjunto supervisado. Cada muestra contiene una ventana histórica y sus valores futuros asociados. Si se utiliza una ventana de longitud L , la entrada puede expresarse como:

$$\mathbf{X}_t = Z_{\{t-L+1\}}, Z_{\{t-L+2\}}, \dots, Z_t$$

donde Z_t incluye el ITCER y las variables explicativas disponibles en el período t .

La salida correspondiente es:

$$\mathbf{Y}_t = y_{\{t+1\}}, y_{\{t+3\}}, y_{\{t+6\}}$$

Por tanto, cada observación de entrenamiento adopta la siguiente forma:

$$(\mathbf{X}_t, \mathbf{Y}_t)$$

2.2.4.10 Evaluación por horizonte

El desempeño de un modelo multi-horizonte debe analizarse por separado para cada horizonte. Un modelo puede presentar un error reducido a un mes y perder precisión a seis meses. Por esta razón, se calculan métricas específicas:

$$\begin{array}{ccc}
MAE_{\{h1\}}, & MAE_{\{h3\}}, & MAE_{\{h6\}} \\
RMSE_{\{h1\}}, & RMSE_{\{h3\}}, & RMSE_{\{h6\}} \\
sMAPE_{\{h1\}}, & sMAPE_{\{h3\}}, & sMAPE_{\{h6\}}
\end{array}$$

También puede calcularse un promedio general:

$$MAE_{promedio} = \frac{MAE_{h1} + MAE_{h3} + MAE_{h6}}{3}$$

2.2.4.11 Modelo ingenuo

El modelo ingenuo de persistencia constituye una referencia básica para evaluar el desempeño de modelos de mayor complejidad. Su supuesto central consiste en considerar que el valor futuro de la variable será igual a su último valor observado.

$$\hat{y}_{\{t+h\}} = y_t$$

para: $h \in 1,3,6$.

Por ejemplo, si el último valor disponible del índice es 102.5, el modelo asignará el mismo valor a los tres horizontes:

$$\hat{y}_{\{t+1\}} = 102.5, \quad \hat{y}_{\{t+3\}} = 102.5, \quad \hat{y}_{\{t+6\}} = 102.5$$

Este modelo no incorpora relaciones complejas ni variables explicativas. Su función principal es establecer una referencia de desempeño. En consecuencia, un modelo más sofisticado debería demostrar una mejora respecto de esta línea base para justificar la incorporación de una mayor complejidad metodológica.

2.2.4.12 Modelo Seasonal Naive (Ingenuo Estacional)

El modelo Seasonal Naive constituye una extensión del modelo ingenuo que considera la estacionalidad de la serie. Su principio consiste en utilizar como pronóstico el valor observado en el mismo período del ciclo estacional anterior.

La ecuación es:

$$y_{\{t+h\}} = y_{\{t+h\}} - m$$

Donde m es el período estacional. Para datos mensuales: $m = 12$

Por ejemplo, para pronosticar el valor correspondiente a enero, el modelo utiliza como referencia el valor observado en enero del año anterior. Este enfoque constituye una línea base particularmente útil cuando la serie presenta patrones estacionales, debido a que permite determinar si un modelo más complejo logra superar un pronóstico basado exclusivamente en la repetición del patrón observado en el ciclo anterior.

En el caso del ITCER mensual, el modelo Seasonal Naive puede utilizarse como referencia adicional al modelo ingenuo de persistencia. La comparación entre ambas líneas base permite evaluar si los modelos propuestos aportan mejoras frente a predicciones basadas únicamente en la persistencia del nivel o en la repetición de patrones estacionales.

2.2.5 Índice de Tipo de Cambio Efectivo Real (ITCER)

2.2.5.1 Definición del ITCER

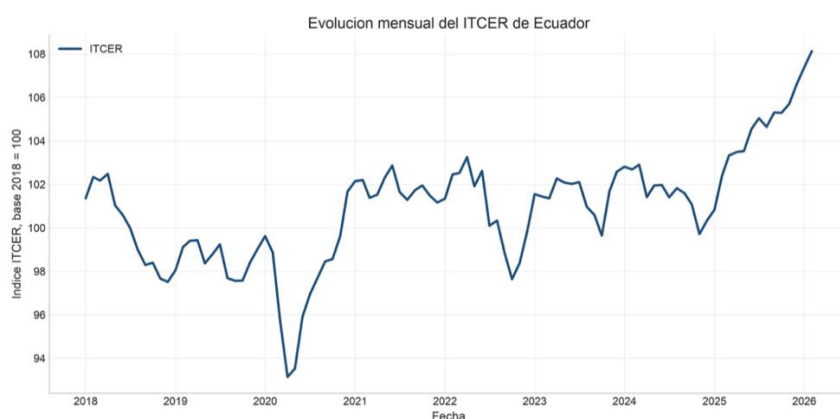
El tipo de cambio real es un indicador que permite comparar el precio relativo de los bienes y servicios de una economía con respecto a los de sus principales socios comerciales, considerando tanto las variaciones de los tipos de cambio nominales como los niveles de precios. Cuando esta comparación se realiza frente a un conjunto de países y se asignan ponderaciones de acuerdo con la importancia de cada socio en el comercio exterior, se obtiene el Índice de Tipo de Cambio Efectivo Real (ITCER).

El Banco Central del Ecuador (BCE) define el ITCER como un indicador que relaciona el tipo de cambio nominal con los niveles relativos de precios internos y externos, considerando las variaciones de los tipos de cambio y de los precios de los países que mantienen relaciones comerciales con Ecuador. En su metodología vigente, el indicador utiliza como período base el año 2018, por lo que su valor se expresa con una base de 2018 = 100.

A diferencia de un tipo de cambio real bilateral, el ITCER no establece una comparación únicamente con un país, sino con una canasta de socios comerciales. La metodología utilizada por el BCE considera 16 países, que en el año base representaron aproximadamente el 80,5 % del comercio exterior ecuatoriano. Entre los principales socios considerados se encuentran Estados Unidos, China, Colombia, México, Perú, Alemania, Panamá, España, Japón y Brasil, entre otros. Las ponderaciones asignadas a cada país se determinan a partir de su participación en el comercio exterior de Ecuador, considerando las exportaciones no petroleras y las importaciones valoradas en términos FOB.

El valor de 100 representa el nivel del índice correspondiente al período base. Por tanto, un valor superior o inferior a 100 refleja una variación relativa respecto de dicho período, pero no permite determinar, por sí solo, si el tipo de cambio se encuentra en un nivel de equilibrio, sobrevaloración o subvaloración. En este sentido, el ITCER debe interpretarse como un indicador de variación relativa del tipo de cambio real respecto de un período de referencia y no como una medida directa del nivel de equilibrio cambiario.

Figura 1
Evolución mensual ITCER



Fuente: Banco Central del Ecuador, 2026

La Figura 1 muestra la evolución mensual del ITCER entre 2018 y 2026. Durante 2018 y 2019, el índice presentó fluctuaciones moderadas alrededor del nivel de referencia. En 2020 se observa una disminución considerable, seguida de una recuperación durante 2021. Posteriormente, entre 2022 y 2024, el indicador mantuvo un comportamiento fluctuante, con períodos alternados de apreciación y depreciación real.

A partir de 2025 se evidencia una tendencia ascendente más marcada, alcanzando valores cercanos a 108 al final del período. De acuerdo con la convención utilizada por el Banco Central del Ecuador, este incremento representa una depreciación real. En términos generales, la trayectoria observada evidencia que el ITCER presenta variaciones a lo largo del tiempo, asociadas a cambios en los precios relativos y en las condiciones cambiarias de Ecuador y sus principales socios comerciales.

2.2.5.2 Cálculo general

De acuerdo con la metodología del Banco Central del Ecuador, el ITCER se calcula mediante la siguiente expresión:

$$ITCER_t = \frac{\prod_{i=1}^n (P_{i,t})^{a_i}}{\frac{P_t}{E_t}} \times 100$$

Donde,

- P_t : Representa el Índice de Precios al Consumidor de Ecuador en el periodo t ;
- $P_{i,t}$: Corresponde al Índice de Precios al Consumidor del socio comercial i ;
- E_t : Representa el componente del tipo de cambio nominal;
- a_i : Es la ponderación asignada al país i , de acuerdo con su participación en el comercio exterior ecuatoriano; y
- n : es el número de países incluidos en la canasta.

La fórmula combina el componente de precios con el componente cambiario. El primero relaciona el nivel de precios de los socios comerciales con el nivel de precios interno, mientras que el segundo incorpora las variaciones de los tipos de cambio de las monedas de los países considerados frente al dólar estadounidense.

Las ponderaciones permiten que los países con una mayor participación comercial tengan una influencia más alta en el índice. Estas se obtienen a partir de la participación de las exportaciones no petroleras y las importaciones de cada país dentro del total considerado. El petróleo se excluye del cálculo de las ponderaciones para evitar que la concentración de las exportaciones petroleras sobredimensione la importancia de determinados socios comerciales.

Según la convención utilizada por el Banco Central del Ecuador, un incremento del ITCER representa una depreciación real, mientras que una disminución indica una apreciación real. Una depreciación real supone que los productos ecuatorianos se vuelven relativamente menos costosos frente a los productos extranjeros. En cambio, una apreciación real indica que los bienes ecuatorianos se encarecen en términos relativos.

2.2.5.3 ITCER en una economía dolarizada

El análisis del ITCER tiene una importancia particular en Ecuador debido a que el país adoptó el dólar estadounidense como moneda de curso legal en el año 2000. Bajo este régimen, Ecuador no dispone de una moneda nacional cuyo tipo de cambio pueda modificarse mediante una devaluación o revaluación propia.

Es decir, la variación del ITCER ecuatoriano depende principalmente de dos factores: los cambios de las monedas de los socios comerciales frente al dólar y el diferencial entre la inflación de Ecuador y la inflación de dichos países. El Banco Central del Ecuador señala que, al estar dolarizado, los movimientos del ITCER son resultado de las devaluaciones o

revaluaciones de las monedas de los socios comerciales, junto con las diferencias de inflación.

Por ejemplo, cuando el dólar se aprecia frente al peso colombiano, al sol peruano o al euro, los productos ecuatorianos pueden encarecerse relativamente en esos mercados. Bajo la convención del BCE, este comportamiento tiende a reflejarse en una disminución del ITCER, es decir, en una apreciación real. Por el contrario, si las monedas de los socios comerciales se fortalecen frente al dólar, los productos ecuatorianos pueden volverse relativamente más accesibles y el ITCER puede aumentar.

La inflación también desempeña un papel central. Si los precios en Ecuador aumentan más rápidamente que los precios externos, se produce una apreciación real, debido a que los bienes nacionales se encarecen en términos relativos. Si la inflación de los socios comerciales es superior a la ecuatoriana, puede producirse una depreciación real y una mejora relativa de la competitividad de precios.

2.2.5.4 Factores macroeconómicos relacionados con el ITCER

En esta investigación se consideran como variables explicativas el IPC de Ecuador, el IPC de Estados Unidos, el índice nominal efectivo del dólar, el precio del petróleo WTI y el EMBI. No todas estas variables forman parte directa de la fórmula oficial del ITCER. El IPC de Ecuador y los índices de precios externos sí intervienen en su cálculo, mientras que el WTI y el EMBI representan condiciones macroeconómicas que podrían aportar información adicional para el pronóstico.

IPC de Ecuador: El Índice de Precios al Consumidor de Ecuador constituye un componente directo del ITCER y representa el nivel de precios doméstico.

IPC de Estados Unidos: Estados Unidos forma parte de la canasta de socios comerciales utilizada por el BCE. Su IPC representa uno de los índices de precios externos incluidos en el cálculo. Debido a que Ecuador y Estados Unidos utilizan la misma moneda, la

relación bilateral no presenta variaciones nominales entre el dólar ecuatoriano y el dólar estadounidense. Por ello, los cambios reales entre ambos países se explican principalmente por el diferencial de inflación.

USD NEER: El Nominal Effective Exchange Rate del dólar es un índice que resume el comportamiento nominal del dólar frente a una canasta de monedas, utilizando ponderaciones comerciales. Esta variable se relaciona con el ITCER ecuatoriano porque Ecuador utiliza el dólar como moneda oficial. Una apreciación del USD NEER indica un fortalecimiento general del dólar frente a otras monedas. En principio, dicho fortalecimiento puede producir una apreciación real de Ecuador y una reducción del ITCER calculado bajo la convención del BCE.

Precio del petróleo WTI: El WTI es uno de los principales precios internacionales de referencia para el petróleo. El BCE lo identifica como el precio de un petróleo ligero extraído principalmente en Texas y utilizado como referencia para la fijación de precios de productos petroleros a nivel internacional. El WTI no forma parte directa de la fórmula del ITCER. Además, el petróleo se excluye de las exportaciones utilizadas para calcular las ponderaciones comerciales del índice. No obstante, puede incorporarse como variable externa debido a su relación con las condiciones económicas de Ecuador, los ingresos petroleros, la actividad económica y determinados costos internos.

EMBI: El EMBI es un indicador relacionado con el riesgo soberano y las condiciones de financiamiento de las economías emergentes. El BCE define el riesgo país como una medida del riesgo asociado con invertir o prestar recursos a un país, considerando la posibilidad de incumplimiento de sus obligaciones financieras o de una crisis económica.

Al igual que el WTI, el EMBI no constituye un componente directo del cálculo del ITCER. Su inclusión se justifica como una variable de contexto financiero que puede reflejar

episodios de incertidumbre, cambios en las condiciones de financiamiento y perturbaciones macroeconómicas.

2.2.6 Modelos estadísticos lineales y econométricos de series temporales

2.2.6.1 Modelos Autorregresivos Integrados de Media Móvil (ARIMA)

El modelo ARIMA (AutoRegressive Integrated Moving Average) es una de las metodologías más empleadas para el análisis y pronóstico de series temporales. Su principal objetivo es describir la dinámica de una serie utilizando la información contenida en sus observaciones pasadas y en los errores de predicción, permitiendo realizar pronósticos cuando la serie presenta dependencia temporal. (Shumway & Stoffer, 2021).

El modelo se representa mediante la notación ARIMA(p,d,q), donde:

p: orden del componente autorregresivo (AR).

d: número de diferenciaciones aplicadas para lograr la estacionariedad.

q: orden del componente de media móvil (MA).

La representación matemática general del modelo es:

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t$$

Esta ecuación indica que el comportamiento de la serie temporal está determinado por un polinomio autorregresivo $\Phi(B)$, un operador de diferenciación $(1 - B)^d$, un polinomio de media móvil $\Theta(B)$ y un término de error aleatorio ε_t . El operador de rezago B desplaza la serie un período hacia atrás, permitiendo expresar el modelo de forma compacta.

2.2.6.2 El componente Autorregresivo (AR-p) y operadores de rezago

El componente autorregresivo (AR) establece que el valor actual de una serie temporal depende linealmente de sus valores observados en periodos anteriores. El parámetro p representa el número de rezagos utilizados para explicar el comportamiento de la serie.

Cuando existe una fuerte dependencia temporal, el modelo AR permite capturar la persistencia de la información histórica y mejorar la capacidad predictiva.

Ecuación general:

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t$$

Donde: X_t es el valor de la serie en el instante t ; c es una constante; ϕ_i son los coeficientes autorregresivos; ϵ_t es un error aleatorio con media cero y varianza constante.

Para simplificar la notación se utiliza el operador de rezago B , definido como:

$$BX_t = X_{t-1}$$

Con este operador, el modelo puede expresarse como:

$$\Phi(B)X_t = \epsilon_t$$

El operador de rezago facilita la representación matemática de modelos de series temporales y permite describir de forma compacta la dependencia entre observaciones consecutivas (Hyndman & Athanasopoulos, 2021; Shumway & Stoffer, 2021).

2.2.6.3 El componente de Media Móvil (MA-q) y choques estocásticos

El componente de media móvil (MA) modela el efecto de perturbaciones aleatorias ocurridas en periodos anteriores. El parámetro q indica el número de errores rezagados incluidos en el modelo. Este componente resulta útil cuando la serie presenta variaciones temporales ocasionadas por eventos externos o choques estocásticos.

Ecuación general:

$$X_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

Donde μ representa la media de la serie; θ_i son los coeficientes de media móvil y ϵ_t corresponde al ruido blanco. Los choques estocásticos representan eventos impredecibles cuyo efecto disminuye conforme transcurre el tiempo (Hyndman & Athanasopoulos, 2021).

2.2.6.4 El proceso de Integración (I-d) y eliminación de no estacionariedad

El componente de integración transforma una serie no estacionaria en estacionaria mediante diferenciación. El parámetro d indica el número de diferencias aplicadas hasta estabilizar la media y la varianza.

Primera diferencia:

$$\nabla X_t = X_t - X_{t-1}$$

Segunda diferencia:

$$\nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

Una serie estacionaria presenta propiedades estadísticas aproximadamente constantes en el tiempo, condición necesaria para estimar correctamente un modelo ARIMA. La prueba Dickey–Fuller aumentada (ADF) suele emplearse para verificar esta condición (Brownlee, 2020; Hyndman & Athanasopoulos, 2021).

2.2.6.5 Metodología de Box-Jenkins (Identificación, estimación y diagnóstico de residuos)

La metodología Box–Jenkins constituye un procedimiento sistemático para la construcción de modelos ARIMA. Comprende varias etapas. En primer lugar, se realiza la identificación del modelo, mediante el análisis de la estacionariedad de la serie, utilizando pruebas como la de Dickey-Fuller aumentada (ADF), así como el estudio de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), con el propósito de seleccionar los valores de (p) , (d) y (q) .

Posteriormente, se lleva a cabo la estimación de los parámetros del modelo mediante métodos como la máxima verosimilitud. A continuación, se realiza el diagnóstico, en el cual se verifica que los residuos presenten un comportamiento compatible con ruido blanco, mediante el análisis de gráficos de residuos, la ACF residual y pruebas estadísticas como la de Ljung–Box.

Finalmente, una vez validado el modelo, se procede a generar los pronósticos para periodos futuros junto con sus respectivos intervalos de confianza. En conjunto, esta metodología busca obtener modelos parsimoniosos, cuyos residuos no presenten autocorrelación y que mantengan una adecuada capacidad predictiva (Hyndman & Athanasopoulos, 2021; Shumway & Stoffer, 2021).

2.2.6.6 Modelos SARIMAX con componentes estacionales y variables exógenas

El modelo SARIMAX es ampliamente utilizado debido a que combina la capacidad de modelar la dependencia temporal, los patrones estacionales y la influencia de variables explicativas externas en un único marco estadístico. Esta integración resulta particularmente adecuada para el pronóstico de indicadores macroeconómicos, donde el comportamiento de una variable suele depender tanto de su propia historia como de factores económicos exógenos que evolucionan simultáneamente (Springer Nature, 2024).

2.2.6.7 Extensión estacional (SARIMA): Especificación matemática de los órdenes (P,D,Q)_s

El modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) amplía el modelo ARIMA al incorporar componentes estacionales que permiten modelar patrones repetitivos presentes en una serie temporal. Mientras que ARIMA describe la dependencia temporal de corto plazo, SARIMA añade parámetros que representan la dinámica estacional mediante un período s , como 12 para datos mensuales o 4 para datos trimestrales (MathWorks, 2024).

La notación general del modelo es:

$$SARIMA(p, d, q)(P, D, Q)_s$$

Donde:

- P : orden autorregresivo estacional.

- D : número de diferenciaciones estacionales.
- Q : orden de medias móviles estacionales.
- s : longitud del período estacional.

La diferenciación estacional se expresa como:

$$\nabla_s = (1 - B^s)$$

Y para un orden de diferenciación D :

$$\nabla_s^D = (1 - B^s)^D$$

La ecuación general del modelo SARIMA es:

$$\Phi(B^s)\varphi(B)\nabla^d\nabla_s^D X_t = \theta(B^s)\theta(B)\varepsilon_t$$

Donde $\Phi(B)$ y $\theta(B)$ representan los componentes autorregresivos y de medias móviles no estacionales, mientras que $\Phi(B^s)$ y $\theta(B^s)$ corresponden a los componentes estacionales.

2.2.6.8 Incorporación de covariables externas (Componente X)

El modelo SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) incorpora variables explicativas externas al modelo SARIMA para mejorar la capacidad predictiva cuando la variable de interés depende de factores externos, como inflación, tasas de interés o tipo de cambio (MathWorks, 2024).

La formulación general es:

$$\Phi(B^s)\varphi(B)\nabla^d\nabla_s^D Y_t = \beta X_t + \theta(B^s)\theta(B)\varepsilon_t$$

Donde:

- Y_t : variable dependiente.
- X_t : vector de variables exógenas.
- β : coeficientes de las variables explicativas.
- ε_t : término de error.

El término βX_t actúa como un componente de regresión lineal incorporado dentro del modelo de series temporales. Cada coeficiente β_i cuantifica el efecto marginal que ejerce la correspondiente variable exógena sobre la variable objetivo, una vez controlados los efectos autorregresivos, estacionales y de medias móviles presentes en la serie.

En varias aplicaciones económicas, el impacto de una variable externa no ocurre de manera inmediata, sino con cierto retraso temporal. Este comportamiento se conoce como dinámica de transferencia y puede modelarse mediante rezagos de las variables exógenas. Matemáticamente, esta relación puede expresarse como:

$$Y_t = \sum_{i=0}^k \beta_i X_{t-i} + N_t$$

Donde:

- X_{t-i} representa la variable externa rezagada i períodos.
- β_i mide el efecto asociado a cada rezago.
- N_t representa la estructura SARIMA que modela la dependencia temporal residual.

Este enfoque resulta especialmente útil en variables macroeconómicas, ya que muchos indicadores producen efectos diferidos. Por ejemplo, modificaciones en las tasas de interés, el tipo de cambio nominal efectivo (NEER), la inflación o los precios internacionales del petróleo pueden impactar el índice de tipo de cambio real únicamente después de varios meses.

Para que las estimaciones del modelo sean consistentes, las variables externas deben satisfacer el supuesto de exogeneidad. Este supuesto establece que las covariables no deben estar correlacionadas con el término de error del modelo, es decir,

$$\text{Cov}(X_t, \varepsilon_t) = 0$$

Cuando esta condición se cumple, los coeficientes estimados representan adecuadamente la influencia de las variables exógenas sobre la variable dependiente. En caso contrario, la presencia de endogeneidad puede generar estimadores sesgados e inconsistentes, reduciendo la capacidad explicativa y predictiva del modelo (MathWorks, 2024).

2.2.7 Aprendizaje automático aplicado al pronóstico económico

El aprendizaje automático, también denominado machine learning, comprende un conjunto de métodos que permiten identificar patrones y estimar relaciones a partir de datos. En lugar de definir de forma previa una ecuación específica para representar el comportamiento de una variable, estos métodos ajustan una función mediante un proceso de entrenamiento. Su propósito es utilizar la información disponible para realizar predicciones sobre observaciones que no fueron empleadas durante dicho proceso (James et al., 2021).

En el pronóstico económico, el aprendizaje automático permite combinar los valores históricos de una variable con otros indicadores que pueden aportar información sobre su comportamiento futuro. Una de sus características es la capacidad de representar relaciones no lineales e interacciones entre variables. Esto puede ser útil cuando el efecto de un indicador cambia según el periodo, el nivel de otras variables o las condiciones generales de la economía.

Bolhuis y Rayner (2020) señalan que estos métodos pueden complementar a los modelos econométricos tradicionales al incorporar varios predictores, relaciones no lineales y procedimientos de validación fuera de la muestra. De manera similar, Goulet Coulombe et al. (2022) encontraron que una parte importante de la utilidad del aprendizaje automático en el pronóstico macroeconómico se relaciona con su capacidad para representar no linealidades, especialmente en periodos de incertidumbre económica y tensión financiera.

No obstante, su aplicación a series económicas requiere considerar que las observaciones mantienen un orden temporal. Los datos futuros no deben utilizarse para entrenar un modelo destinado a pronosticar periodos anteriores. Por esta razón, la preparación

de las variables, la división de la muestra y la evaluación del modelo deben respetar la secuencia cronológica de los datos.

2.2.8 Aprendizaje supervisado

El aprendizaje supervisado es un modelo que aprende a partir de ejemplos históricos compuestos por variables de entrada y una salida conocida. Su objetivo es identificar patrones y relaciones que permitan realizar predicciones sobre nuevos datos. En el ámbito económico, este enfoque ha ganado relevancia debido a su capacidad para incorporar múltiples variables explicativas y modelar relaciones que no siempre pueden ser representadas adecuadamente mediante métodos estadísticos tradicionales (James et al., 2021; Bolhuis & Rayner, 2020).

El aprendizaje supervisado se basa en observaciones para las cuales se conocen tanto las variables de entrada como la variable de salida. El modelo utiliza estos datos para aprender una función que relacione ambas partes:

$$y_i = f(X_i) + \varepsilon_i$$

donde (y_i) representa la variable que se desea estimar, (X_i) contiene las variables explicativas, (f) es la función aprendida por el modelo y (ε_i) corresponde al componente que no puede explicarse con la información disponible.

Durante el entrenamiento, el algoritmo genera una predicción $(\hat{y}_i)$ y la compara con el valor real (y_i) . Esta diferencia se mide mediante una función de pérdida. En problemas de regresión puede utilizarse, por ejemplo, el error cuadrático medio:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde (n) es el número de observaciones. El proceso de entrenamiento busca encontrar una función que reduzca el valor de esta pérdida.

Para aplicar aprendizaje supervisado a una serie temporal es necesario transformar la secuencia original en un conjunto de entradas y salidas. Las variables de entrada pueden estar

formadas por valores anteriores de la variable objetivo, conocidos como rezagos, y por valores históricos de otras variables explicativas.

En el caso del ITCER, un vector de entrada podría representarse de la siguiente manera:

$$X_t = [ITCER_{t-1}, ITCER_{t-2}, IPC_{EC,t-1}, IPC_{US,t-1}, NEER_{t-1}, WTI_{t-1}, EMBI_{t-1}]$$

La variable de salida correspondería al valor futuro que se desea pronosticar:

$$y_t = ITCER_{t+h}$$

donde (h) representa el horizonte de pronóstico. Para estimar el ITCER a uno, tres y seis meses pueden definirse tres salidas:

$$y_t = [ITCER_{t+1}, ITCER_{t+3}, ITCER_{t+6}]$$

De esta forma, cada observación contiene información disponible hasta el periodo (t), mientras que la salida corresponde a uno o varios periodos posteriores.

Además de los rezagos, pueden incluirse variables derivadas, como variaciones mensuales, diferencias interanuales, medias móviles, desviaciones estándar móviles o valores máximos y mínimos de periodos anteriores. Estas características permiten resumir el comportamiento reciente de la serie y proporcionar información adicional al modelo.

Las variables derivadas deben calcularse utilizando únicamente datos conocidos hasta el momento del pronóstico. Cuando una variable de entrada contiene información proveniente de periodos futuros se produce una fuga de información o data leakage. Este problema genera resultados de evaluación demasiado favorables, pero que no pueden reproducirse al realizar pronósticos reales.

2.2.8.1 Validación temporal

La evaluación de los modelos de aprendizaje automático suele realizarse mediante la división de los datos en conjuntos de entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para estimar los parámetros del modelo, el conjunto de validación

permite seleccionar hiperparámetros y el conjunto de prueba se reserva para la evaluación final.

En datos independientes es frecuente dividir las observaciones de manera aleatoria. Sin embargo, este procedimiento no es adecuado para series temporales, debido a que puede colocar observaciones futuras en el conjunto de entrenamiento y observaciones anteriores en el conjunto de prueba.

En series temporales debe mantenerse el orden cronológico. Una alternativa consiste en utilizar las observaciones iniciales para entrenamiento y las más recientes para prueba. También puede aplicarse una validación de origen móvil, conocida como rolling origin, en la cual el modelo se entrena varias veces utilizando ventanas que avanzan a través del tiempo.

En una ventana expansiva, el conjunto de entrenamiento aumenta en cada iteración:

$$[1, \dots, t_1 \rightarrow 1, \dots, t_2 \rightarrow 1, \dots, t_3]$$

En cada etapa, el modelo se evalúa con observaciones posteriores al conjunto utilizado para entrenarlo. Este procedimiento permite estimar de manera más realista cómo habría funcionado el modelo en diferentes momentos históricos. Hyndman y Athanasopoulos (2021) indican que, en la validación temporal, cada conjunto de entrenamiento debe contener únicamente observaciones anteriores al periodo empleado como prueba.

2.2.8.2 Modelos basados en árboles de decisión

Los árboles de decisión son modelos que dividen el espacio de las variables mediante una secuencia de reglas. Cada regla utiliza una variable explicativa y un punto de corte para separar las observaciones en grupos con comportamientos similares.

Una división podría expresarse como:

$$USD\ NEER_{t-1} < c$$

donde (c) es un punto de corte seleccionado durante el entrenamiento. Las observaciones que cumplen esta condición se asignan a una rama del árbol y las restantes se ubican en otra.

En cada nodo, el algoritmo evalúa distintas variables y posibles puntos de corte. En un problema de regresión, se selecciona la división que produzca la mayor reducción del error dentro de los grupos resultantes. Este proceso se repite hasta alcanzar un criterio de parada, como una profundidad máxima, una cantidad mínima de observaciones por nodo o una mejora mínima en la función de pérdida.

Las divisiones generan regiones del espacio de entrada. La predicción para una observación se obtiene a partir del nodo terminal en el que esta se ubica. En regresión, normalmente se utiliza el promedio de los valores observados dentro de dicho nodo:

$$\hat{y}(X) = \frac{1}{N_m} \sum_{i \in R_m} y_i$$

Donde R_m representa una región terminal y N_m es el número de observaciones contenidas en ella.

Una ventaja de los árboles es que no requieren asumir una relación lineal entre los predictores y la variable objetivo. También pueden identificar interacciones de manera automática. Por ejemplo, el árbol puede realizar primero una división según el comportamiento del dólar y posteriormente aplicar reglas diferentes según el IPC de Ecuador o el nivel del EMBI.

Los árboles tampoco requieren normalmente que las variables se encuentren en la misma escala. Esto se debe a que sus decisiones se basan en puntos de corte y no en distancias entre observaciones. Por tanto, la normalización no suele ser un requisito para este tipo de modelos.

Sin embargo, un árbol individual puede presentar alta variabilidad. Un pequeño cambio en los datos de entrenamiento puede modificar los puntos de corte y producir una estructura diferente. Además, un árbol demasiado profundo puede generar regiones con pocas observaciones y ajustarse en exceso a la muestra.

Para reducir estas limitaciones se utilizan métodos de ensamble, los cuales combinan varios árboles. La idea general es que la predicción conjunta puede ser más estable y precisa que la obtenida mediante un único árbol.

Uno de estos métodos es *Random Forest*. Este algoritmo construye múltiples árboles a partir de muestras de los datos y subconjuntos aleatorios de variables. En problemas de regresión, la predicción final se obtiene mediante el promedio de los resultados de los árboles. Breiman (2001) plantea que esta aleatorización permite reducir la correlación entre los árboles y mejorar la capacidad de generalización del conjunto.

2.2.8.3 Gradient Boosting

El Gradient Boosting es otro método de ensamble basado en árboles, pero su forma de entrenamiento es diferente. En lugar de construir árboles independientes y promediar sus resultados, los árboles se agregan de manera secuencial. Cada nuevo árbol se entrena para corregir parte de los errores producidos por el modelo acumulado.

El procedimiento comienza con una predicción inicial. Cuando se utiliza el error cuadrático, esta predicción puede corresponder al promedio de la variable objetivo:

$$F_0(X) = \bar{y}$$

Posteriormente, se calculan los errores o residuos de esta primera estimación:

$$r_i = y_i - F_0(X_i)$$

El primer árbol se entrena utilizando estos residuos como variable de salida. Su resultado se incorpora a la predicción inicial:

$$F_1(X) = F_0(X) + \eta f_1(X)$$

Donde, $f_1(X)$ representa el árbol incorporado y η corresponde a la tasa de aprendizaje.

El proceso continúa de manera secuencial. En la iteración (m), se añade un nuevo árbol al modelo anterior:

$$F_m(X) = F_{m-1}(X) + \eta f_m(X)$$

Friedman (2001) formuló este procedimiento como una optimización en el espacio de funciones. En cada iteración, el nuevo árbol busca aproximar el gradiente negativo de la función de pérdida. Dicho gradiente indica la dirección en la que debe modificarse el modelo para reducir el error.

Cuando la función de pérdida es el error cuadrático, el gradiente negativo coincide con los residuos entre los valores reales y las predicciones. Para otras funciones de pérdida, el algoritmo utiliza valores derivados de sus respectivas pendientes.

La predicción final puede representarse como:

$$\hat{y}(X) = F_0(X) + \eta \sum_{m=1}^M f_m(X)$$

Donde (M) representa el número total de árboles.

La tasa de aprendizaje (η) controla el aporte de cada árbol. Un valor pequeño hace que las correcciones sean más moderadas y suele requerir un número mayor de iteraciones. Un valor alto permite un ajuste más rápido, pero puede aumentar la posibilidad de sobreajuste.

La profundidad de los árboles determina el nivel de interacción que puede representar el modelo. Los árboles poco profundos capturan relaciones más simples, mientras que árboles más profundos permiten representar interacciones más complejas. Sin embargo, también incrementan la complejidad del modelo.

En el pronóstico económico, Gradient Boosting puede combinar rezagos de la variable objetivo, indicadores macroeconómicos y características derivadas. Su capacidad

para representar interacciones y relaciones no lineales puede ser útil cuando la dinámica de una variable cambia entre distintos periodos.

A partir de este enfoque se han desarrollado diferentes implementaciones de árboles potenciados. Una de ellas es LightGBM, que utiliza árboles de decisión y procedimientos orientados a mejorar la eficiencia del entrenamiento. Ke et al. (2017) propusieron esta implementación mediante técnicas como el muestreo basado en la magnitud del gradiente y la agrupación de variables excluyentes. Sus características serán desarrolladas en el siguiente apartado.

En esta investigación, LightGBM se emplea como modelo de aprendizaje automático para pronosticar el ITCER a uno, tres y seis meses. Sus resultados se utilizan como referencia para comparar el desempeño del modelo basado en aprendizaje profundo.

2.2.7 LightGBM

2.2.7.5 Definición

LightGBM, abreviatura de Light Gradient Boosting Machine, es un algoritmo de aprendizaje automático basado en árboles de decisión potenciados por gradiente. Fue desarrollado por Ke et al. (2017) como una implementación eficiente de los modelos Gradient Boosting Decision Trees o GBDT.

El algoritmo puede utilizarse en tareas de regresión, clasificación y ordenamiento. En un problema de regresión, como el pronóstico del Índice de Tipo de Cambio Efectivo Real, LightGBM estima un valor numérico continuo a partir de un conjunto de variables explicativas.

LightGBM es un modelo de ensamble. Esto significa que la predicción final no se obtiene mediante un único árbol, sino mediante la combinación de varios árboles construidos de forma secuencial. Cada árbol nuevo intenta reducir los errores que todavía presenta el conjunto de árboles anteriores.

La implementación utiliza árboles basados en histogramas y un crecimiento por hojas denominado leaf-wise. También dispone de mecanismos de regularización, muestreo de observaciones, selección parcial de variables y detención anticipada del entrenamiento. Estas características permiten controlar la complejidad y reducir el costo computacional del modelo.

LightGBM no es un modelo temporal por sí mismo. El algoritmo no conoce directamente el orden cronológico de las observaciones ni conserva una memoria de los periodos anteriores. Para aplicarlo al pronóstico de series temporales, es necesario transformar la serie en un conjunto de variables de entrada, mediante rezagos, diferencias, variaciones y otras características calculadas con información pasada.

La inclusión de modelos basados en árboles de decisión potenciados por gradiente, como es el caso de LightGBM, cumple una función metodológica al establecer un punto de comparación o benchmark no lineal intermedio entre la econometría clásica y el aprendizaje profundo. Si bien las innovaciones de esta arquitectura se centran en la eficiencia computacional para el manejo de macrodatos mediante el muestreo unilateral basado en gradientes y la agrupación de características exclusivas, en conjuntos de datos de escala moderada su valor no reside en la optimización de recursos, sino en su flexibilidad para identificar relaciones no lineales y efectos de umbral entre las variables exógenas sin asumir supuestos paramétricos estrictos (Ke et al., 2017)

2.2.9.2 Gradient Boosting y aprendizaje secuencial

El funcionamiento de LightGBM parte del principio de Gradient Boosting. Este método construye una función predictiva mediante la suma de varios árboles de decisión:

$$F_M(X) = F_0(X) + \eta \sum_{m=1}^M f_m(X)$$

Donde:

- $F_0(X)$: es la predicción inicial;
- $f_m(X)$: representa el árbol construido en la iteración (m);
- M : es el número total de árboles;
- η : es la tasa de aprendizaje.

En un problema de regresión, la predicción inicial puede corresponder al promedio de la variable objetivo. A partir de esta primera estimación, el algoritmo calcula el error de cada observación y construye un árbol para aproximar la dirección en la que debe corregirse el modelo.

La construcción es secuencial porque el árbol (m) depende de los resultados obtenidos por los árboles anteriores. Por tanto, los árboles no se entrenan de manera completamente independiente. Cada nueva etapa trabaja sobre el modelo acumulado:

$$F_m(X) = F_{m-1}(X) + \eta f_m(X)$$

La tasa de aprendizaje (η) reduce el aporte individual de cada árbol. Una tasa pequeña genera correcciones moderadas y suele requerir más iteraciones. Una tasa elevada permite que cada árbol tenga mayor influencia, pero puede hacer que el modelo se ajuste con demasiada rapidez a los datos de entrenamiento.

La estimación de cada árbol se realiza minimizando una función objetivo. De manera general, esta función puede expresarse como:

$$\mathcal{J} = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m)$$

donde (L) es la función de pérdida que mide la diferencia entre el valor real (y_i) y la predicción $(\hat{y}_i)$ mientras que $(\Omega(f_m))$ representa una penalización asociada con la complejidad de los árboles.

En regresión, LightGBM permite utilizar distintas funciones objetivo. Entre ellas se encuentran el error cuadrático, el error absoluto, la pérdida de Huber y la regresión por cuantiles. La función predeterminada para regresión es la pérdida cuadrática o L2.

La pérdida cuadrática asigna una penalización mayor a los errores grandes:

$$L(y_i, \hat{y}_i) = \frac{1}{2}(y_i - \hat{y}_i)^2$$

Esta función puede ser apropiada cuando se desea reducir el RMSE. Sin embargo, también puede hacer que observaciones atípicas tengan mayor influencia. Si la serie presenta valores extremos, pueden evaluarse funciones más robustas, como la pérdida absoluta o la pérdida de Huber.

2.2.9.3 Construcción de árboles mediante histogramas

En un árbol tradicional, cada valor de una variable continua puede evaluarse como posible punto de corte. Este procedimiento puede ser costoso cuando existen muchas observaciones o valores diferentes.

LightGBM utiliza un algoritmo basado en histogramas. Antes de construir los árboles, los valores continuos de cada variable se agrupan en intervalos discretos denominados bins.

Por ejemplo, si el USD NEER contiene numerosos valores diferentes, estos pueden ser agrupados en un conjunto menor de intervalos. Durante la construcción del árbol, el algoritmo no evalúa cada valor original, sino los límites entre dichos intervalos.

Para cada *bin*, LightGBM acumula:

- el número de observaciones;
- la suma de gradientes;
- la suma de Hessianos.

Posteriormente, calcula la ganancia de las posibles divisiones utilizando estas sumas acumuladas. Una vez construido el histograma, el costo de evaluar las divisiones depende principalmente del número de intervalos y no del número total de valores distintos de la variable.

El uso de histogramas reduce el tiempo de entrenamiento y el uso de memoria. Además, LightGBM puede obtener el histograma de una hoja mediante la resta entre el histograma del nodo padre y el de su hoja hermana, evitando recalcular ambas estructuras.

Una cantidad elevada de bins permite representar puntos de corte con mayor precisión, pero aumenta el costo computacional y puede facilitar divisiones muy específicas. Una cantidad reducida genera una discretización más gruesa y puede actuar como una forma adicional de regularización.

Para el caso de este trabajo, con una muestra mensual cercana a 108 registros, el objetivo principal de los histogramas no es reducir un costo computacional elevado. En este caso, una cantidad moderada de intervalos puede ayudar a evitar divisiones construidas a partir de diferencias muy pequeñas entre observaciones.

2.2.9.4 Crecimiento por hojas

Una de las diferencias principales entre LightGBM y otras implementaciones de árboles potenciados es su estrategia de crecimiento. Los algoritmos de crecimiento por niveles, o level-wise, dividen los nodos que se encuentran en una misma profundidad antes de avanzar al siguiente nivel. Como resultado, suelen producir árboles relativamente equilibrados.

LightGBM utiliza un crecimiento leaf-wise o best-first. En cada etapa examina todas las hojas disponibles y divide aquella que genera la mayor reducción de la pérdida. Por ello, una rama puede crecer más que las demás y formar un árbol asimétrico.

Para un número fijo de hojas, esta estrategia suele reducir más rápidamente la pérdida de entrenamiento. Sin embargo, también puede generar ramas profundas que describen grupos muy pequeños de observaciones. Este riesgo es mayor cuando la muestra es reducida. La documentación de LightGBM advierte que el crecimiento por hojas puede producir sobreajuste en conjuntos pequeños si no se limita adecuadamente.

La complejidad se controla principalmente mediante:

- `num_leaves`: número máximo de hojas por árbol;
- `max_depth`: profundidad máxima permitida;
- `min_data_in_leaf`: número mínimo de observaciones por hoja;
- `min_sum_hessian_in_leaf`: suma mínima de Hessianos dentro de una hoja;
- `min_gain_to_split`: reducción mínima de pérdida requerida para dividir.
- `num_leaves` es uno de los parámetros más importantes. Un número alto permite representar relaciones más específicas, pero también aumenta el riesgo de sobreajuste. En árboles leaf-wise, limitar solamente la profundidad puede ser insuficiente, por lo que debe controlarse también la cantidad de hojas.
- `min_data_in_leaf` impide que el algoritmo forme hojas con muy pocas observaciones. En una muestra corta, este parámetro tiene especial importancia. Si una hoja contiene uno o dos meses, la regla aprendida podría representar un episodio particular y no una relación estable.

2.2.8 Aprendizaje profundo para series temporales

2.2.8.5 Redes neuronales artificiales

Las redes neuronales artificiales (Artificial Neural Networks, ANN) son modelos computacionales inspirados en el funcionamiento del sistema nervioso biológico. Están conformadas por un conjunto de neuronas organizadas en capas (entrada, ocultas y salida), las cuales transforman la información mediante operaciones matemáticas con el objetivo de aprender la relación entre las variables de entrada y la variable objetivo.

Su principal ventaja radica en la capacidad de aprender relaciones complejas y no lineales entre variables, característica que ha favorecido su aplicación en problemas de clasificación, regresión y pronóstico de series temporales.

El funcionamiento de una neurona puede representarse mediante la siguiente expresión:

$$y = f(\sum_{i=1}^n w_i x_i + b)$$

Donde:

- x_i : son las variables de entrada.
- w_i : representan los pesos aprendidos durante el entrenamiento.
- b : es el sesgo (bias).
- $f(-)$: corresponde a la función de activación.
- y : es la salida de la neurona.

Durante el entrenamiento, la red ajusta automáticamente los pesos mediante algoritmos de optimización con el fin de minimizar el error entre los valores predichos y los observados.

En el contexto del ITCER, las entradas de la red pueden estar conformadas por variables macroeconómicas como el Índice de Precios al Consumidor (IPC), el riesgo país, el precio del petróleo WTI y otros indicadores económicos, mientras que la salida corresponde

al valor futuro del ITCER. De esta manera, la red aprende las relaciones no lineales existentes entre dichas variables para realizar el pronóstico.

2.2.8.6 Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes fueron (Recurrent Neural Networks, RNN) desarrolladas para trabajar con datos secuenciales, incorporando mecanismos que permiten utilizar información de observaciones anteriores durante el proceso de predicción. Esta característica las hace especialmente adecuadas para el análisis de series temporales, donde los valores pasados influyen en el comportamiento futuro de la variable de interés. Sin embargo, presentan limitaciones para capturar dependencias de largo plazo debido a problemas asociados al entrenamiento de la red.

Su funcionamiento básico puede expresarse como:

$$h_t = f(W_x x_t + W_h h_{t-1} + b)$$

Donde:

- x_t : representa la entrada en el instante t .
- h_{t-1} : corresponde al estado oculto anterior.
- h_t : es el nuevo estado oculto.
- W_x y W_h : son matrices de pesos.
- b : es el término de sesgo.

Gracias a esta estructura, las RNN pueden utilizar información histórica para generar una predicción. Sin embargo, cuando la secuencia es muy larga, presentan dificultades para conservar información relevante debido al problema de desaparición del gradiente (vanishing gradient).

En el caso del ITCER, una RNN puede procesar la evolución mensual del índice considerando que el valor observado en cada período depende, en cierta medida, del comportamiento registrado en los meses anteriores.

2.2.8.7 Long Short-Term Memory (LSTM)

El modelo LSTM (Long Short-Term Memory) es una arquitectura de red neuronal recurrente (RNN). Durante el entrenamiento, el modelo compara sus predicciones con los valores reales y ajusta sus pesos mediante retro propagación (Backpropagation Through Time, BPTT). Ese proceso corresponde al aprendizaje supervisado.

Las redes Long Short-Term Memory (LSTM) son una evolución de las redes recurrentes tradicionales, desarrolladas para capturar dependencias temporales de largo plazo. Su principal característica es la incorporación de un estado de memoria y tres puertas internas que controlan el flujo de información durante el entrenamiento.

El funcionamiento del modelo puede resumirse mediante las siguientes ecuaciones:

Puerta de olvido:

$$f_t = \sigma (W_f [h_{t-1}, x_t] + b_f)$$

Puerta de entrada:

$$i_t = \sigma (W_i [h_{t-1}, x_t] + b_i)$$

Actualización de la memoria:

$$C_t = f_t C_{t-1} + i_t C^{\sim}t$$

Salida:

$$h_t = O_t \tanh(C_t)$$

Estas puertas permiten que la red conserve únicamente la información relevante y descarte aquella que no aporta al proceso de predicción.

Durante el entrenamiento, el modelo compara las predicciones obtenidas con los valores reales y ajusta sus parámetros mediante el algoritmo Backpropagation Through Time (BPTT) utilizando un optimizador como Adam y una función de pérdida, generalmente el Error Cuadrático Medio (MSE).

Para el pronóstico del ITCER, una LSTM recibe como entrada una secuencia histórica de varios meses junto con variables macroeconómicas asociadas. A partir de esta información, aprende la relación temporal existente entre los indicadores y estima el comportamiento futuro del índice.

2.2.8.8 LSTM Multi-horizonte

El enfoque multi-horizonte permite generar predicciones para varios períodos futuros de manera simultánea. A diferencia de los modelos que estiman únicamente el siguiente valor de la serie, una LSTM multi-horizonte produce un conjunto de pronósticos correspondientes a distintos horizontes temporales a partir de una misma secuencia de entrada. Esta característica resulta especialmente útil en problemas económicos, donde la planificación y la toma de decisiones requieren estimaciones para diferentes períodos futuros.

El enfoque multi-horizonte permite estimar varios períodos futuros mediante una única ejecución del modelo. En lugar de generar únicamente el siguiente valor de la serie, la red produce un conjunto de predicciones para diferentes horizontes temporales.

Si la ventana de entrada contiene observaciones, el modelo puede representarse como:

$$\hat{y}_{\{t+1:t+h\}} = f_{\{LSTM\}}(x_{\{t-p+1:t\}}) = (\hat{y}_{\{t+1\}}, \hat{y}_{\{t+2\}}, \dots, \hat{y}_{\{t+h\}})$$

Donde:

- x : representa las variables de entrada.
- h : corresponde al número de horizontes de predicción.

Por ejemplo, utilizando una ventana de los últimos 12 meses, el modelo puede generar simultáneamente el pronóstico del ITCER para los siguientes 1, 3 y 6 meses, evitando que el error acumulado de una predicción afecte a las siguientes.

En esta investigación, la LSTM multi-horizonte recibe como entrada la evolución histórica del ITCER junto con variables macroeconómicas como IPC, riesgo país, WTI y otras variables seleccionadas. La salida del modelo corresponde a los valores esperados del ITCER para varios meses futuros, constituyendo uno de los modelos base de comparación frente al Temporal Fusion Transformer.

2.2.8.9 Temporal Fusion Transformer (TFT)

El Temporal Fusion Transformer (TFT) es una arquitectura de aprendizaje profundo diseñada para el pronóstico multi-horizonte de series temporales. Este modelo combina mecanismos de atención con estructuras recurrentes para integrar información histórica y variables explicativas de distinta naturaleza. Una de sus principales ventajas es la capacidad de identificar la importancia relativa de las variables utilizadas en el proceso de predicción, proporcionando un mayor nivel de interpretabilidad en comparación con otros modelos de aprendizaje profundo (Lim et al., 2021). Debido a estas características, el TFT se ha convertido en una de las arquitecturas más estudiadas para el pronóstico de series temporales multivariadas.

Este mecanismo permite asignar diferentes niveles de importancia a cada instante temporal y a cada variable durante el proceso de predicción, favoreciendo la identificación de patrones complejos presentes en la serie.

En el contexto del ITCER, el modelo puede integrar simultáneamente variables históricas del índice, indicadores macroeconómicos nacionales e internacionales y variables temporales, como el mes o el trimestre del año. Además de generar pronósticos para varios

horizontes futuros, el TFT proporciona medidas de interpretabilidad que permiten identificar cuáles variables tuvieron mayor influencia en cada predicción, facilitando el análisis económico de los resultados y fortaleciendo la toma de decisiones.

2.2.8.10 Interpretabilidad en modelos TFT

La interpretabilidad del TFT se fundamenta en dos mecanismos principales: selección dinámica de variables (Variable Selection Networks) y mecanismo de atención temporal interpretable (Interpretable Multi-Head Attention).

Ambos mecanismos generan pesos internos durante el entrenamiento que pueden analizarse posteriormente para explicar el comportamiento del modelo.

2.2.8.11 Interpretabilidad mediante la selección de variables

Antes de comenzar el procesamiento temporal, el TFT determina automáticamente qué variables son más importantes en cada instante.

Supóngase que existen m variables de entrada:

$$X = \{x_1, x_2, \dots, x_m\}$$

Cada variable se transforma primero en un vector de representación (embedding o transformación lineal). Posteriormente, todas las variables ingresan a una Gated Residual Network (GRN) que calcula un puntaje para cada una.

Finalmente, estos puntajes se normalizan mediante una función Softmax:

$$v_t = \text{Softmax}(GRN(\Xi_t, c_s))$$

Donde

- Ξ_t representa todas las variables en el instante t ,
- c_s corresponde al contexto generado por las variables estáticas,
- v_t contiene los pesos de importancia de cada variable.

2.2.8.12 Interpretabilidad mediante Self-Attention

Una vez seleccionadas las variables, el TFT debe determinar qué momentos del pasado son relevantes para predecir el futuro. Para ello utiliza un mecanismo denominado Interpretable Multi-Head Attention, donde cada observación temporal genera tres vectores: Query (Q), Key (K) y Value (V)

La atención se calcula mediante

$$Attention(Q, K, V) = A(Q, K)V$$

Donde

$$A(Q, K) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)$$

Aquí:

- QK^T : mide la similitud entre dos instantes temporales;
- d : es la dimensión de los vectores;
- $Softmax$: transforma estas similitudes en pesos de atención.

En el Transformer tradicional, cada cabeza (head) utiliza diferentes vectores Value, lo que dificulta interpretar la contribución de cada una.

El TFT modifica esta arquitectura haciendo que todas las cabezas compartan los mismos valores y promediando los pesos de atención:

$$\tilde{H} = \left(\frac{1}{m_H}\right) \sum_{h=1}^{m_H} Attention(QW_Q^h, KW_K^h, VW_V)$$

Donde

- m_H es el número de cabezas de atención;
- W_Q, W_K, W_V son las matrices de transformación.

Gracias a esta modificación es posible analizar un único mapa de atención representativo de todas las cabezas, lo que facilita la interpretación sin sacrificar capacidad de aprendizaje.

Atención temporal

La salida del mecanismo de atención puede escribirse como

$$\beta(t, \tau) = \sum_{n=-k}^T \alpha(t, n, \tau) \tilde{\theta}(t, n)$$

Donde

- $\alpha(t, n, \tau)$: representa el peso de atención asignado al instante n para pronosticar el horizonte τ ;
- $\tilde{\theta}(t, n)$: corresponde a la representación de las características temporales.

El valor de α indica directamente la importancia relativa de cada observación histórica. Si un rezago recibe un peso elevado, significa que esa información fue determinante para el pronóstico.

2.2.8.13 TiDE / Time-series Dense Encoder

El Time-series Dense Encoder (TiDE) es una arquitectura de aprendizaje profundo diseñada para el pronóstico de series temporales a múltiples pasos. Fue propuesta por Das et al. (2023) y se basa principalmente en perceptrones multicapa, conocidos como Multilayer Perceptrons o MLP. Su estructura sigue un esquema de codificador y decodificador, pero no utiliza mecanismos recurrentes, convoluciones ni capas de autoatención.

TiDE surgió a partir de investigaciones que mostraron que modelos lineales relativamente simples podían obtener resultados competitivos frente a arquitecturas Transformer en tareas de pronóstico de largo horizonte. Sin embargo, los modelos lineales presentan dificultades para representar relaciones no lineales y para integrar variables

externas de manera flexible. TiDE busca mantener una estructura computacional sencilla, pero incorpora capas densas no lineales y covariables dinámicas para aumentar su capacidad de representación (Das et al., 2023; Zeng et al., 2023).

El objetivo del modelo es encontrar una función de pronóstico f que asigne de forma precisa el historial, las covariables dinámicas y los atributos estáticos a la predicción del futuro:

$$\widehat{y_{L+1:L+H}^{(i)}} = f_{\theta}(y_{1:L}^{(i)}, x_{1:L+H}^{(i)}, a^{(i)})$$

Donde:

- $y_{1:L}^{(i)}$: representa los valores históricos de la variable objetivo;
- L es la longitud de la ventana de entrada;
- $x_{1:L+H}^{(i)}$: contiene las covariables dinámicas. Las cuales se conocen de antemano tanto para el pasado como para el horizonte futuro (por ejemplo, el día de la semana, días festivos o promociones comerciales)
- $a^{(i)}$: representa los atributos estáticos. Los atributos estáticos de la serie temporal que no cambian con el tiempo (por ejemplo, identificadores o categorías de un producto).
- H : es el horizonte de pronóstico;
- $\hat{y}_{L+1:L+H}$: corresponde a las predicciones futuras.

Esta formulación permite que el modelo genere todos los periodos del horizonte en una sola ejecución, sin utilizar de forma recursiva las predicciones anteriores.

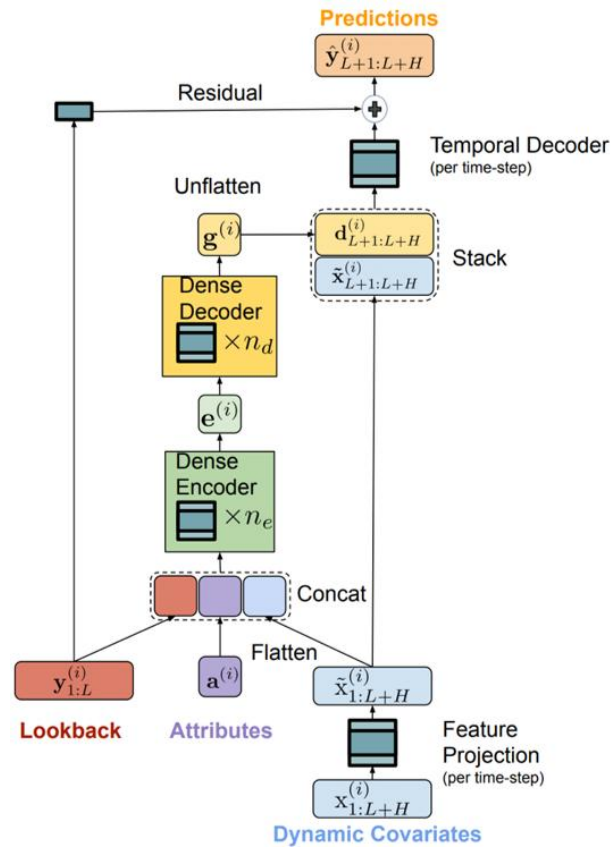
2.2.8.14 Arquitectura del Modelo TiDE

La Figura 2 presenta la arquitectura general del modelo TiDE (Time-series Dense Encoder), en la cual se integran la información histórica, los atributos y las covariables

dinámicas para generar predicciones sobre múltiples horizontes temporales. A continuación, se describen los principales componentes y etapas que conforman esta arquitectura.

Figura 2

Arquitectura del codificador de TiDE



Nota. 1 Tomado de Long-term Forecasting with TiDE: Time-series Dense Encoder (p. 5), por A. Das, W. Kong, A. Leach, S. Mathur, R. Sen y R. Yu, 2024, arXiv

Bloque Residual (Residual Block): Es la unidad fundamental de la arquitectura.

Consiste en un Perceptrón Multicapa (MLP) con una capa oculta y activación ReLU. Cuenta con una conexión de salto (skip connection) puramente lineal, aplica dropout en la capa lineal que mapea la capa oculta hacia la salida, y finalmente implementa normalización de capa (Layer Norm) en su salida.

Proyección de Características (Feature Projection): Dado que aplanar las covariables dinámicas para todo el contexto y horizonte daría como resultado un vector

prohibitivamente grande de tamaño $(L + H)r$, TiDE introduce un paso de reducción de dimensionalidad. Utilizando un bloque residual, mapea las covariables $x_t^{(i)}$ de cada paso temporal de forma independiente a una proyección de menor dimensión $\tilde{r}$ (denominada `temporalWidth`):

$$\widetilde{x_t^{(i)}} = \text{ResidualBlock}(x_t^{(i)})$$

Codificador Denso (Dense Encoder): El codificador recibe un vector unificado creado al apilar y aplanar todas las covariables proyectadas (pasadas y futuras), concatenadas con los atributos estáticos $a^{(i)}$ y el historial de la serie temporal $y_{1:L}^{(i)}$. Esta combinación se mapea mediante múltiples bloques residuales (n_e capas) a una representación incrustada u oculta ($e^{(i)}$) de tamaño *hiddenSize*:

$$e^{(i)} = \text{Encoder} \left(\left[y_{1:L}^{(i)}; \widetilde{x_{1:L+H}^{(i)}}; a^{(i)} \right] \right)$$

Decodificador Denso (Dense Decoder): El decodificador toma la incrustación $e^{(i)}$ y la procesa a través de una pila de varios bloques residuales (n_d capas con tamaño *hiddenSize*). La última capa del decodificador mapea la representación a un vector de tamaño $H \times p$ (donde p es la dimensión `decoderOutputDim`), el cual es posteriormente reestructurado (reshape) en una matriz $D^{(i)} \in \mathbb{R}^{p \times H}$. Cada columna $d_t^{(i)}$ de esta matriz representa el vector decodificado para el periodo t dentro del horizonte de predicción.

Decodificador Temporal (Temporal Decoder): Para generar las predicciones finales, el decodificador temporal (un bloque residual con tamaño de salida igual a 1) procesa de manera independiente para cada paso t del horizonte el vector decodificado $d_t^{(i)}$ concatenado directamente con las covariables proyectadas del futuro $\widetilde{x_{L+t}^{(i)}}$:

$$\widehat{y_{L+t}^{(i)}} = \text{TemporalDecoder} \left(\left[d_t^{(i)}; \widetilde{x_{L+t}^{(i)}} \right] \right) \quad \forall t \in [H]$$

Este diseño añade una "autopista" (highway) desde las covariables futuras directamente hacia la predicción de ese instante temporal específico, evitando que señales directas e inmediatas (como un día festivo sobre la demanda) se atenúen o diluyan durante el proceso de codificación global

Conexión Residual Global: Adicionalmente, TiDE incorpora una conexión residual global que mapea linealmente el historial de retrospectiva $y_{1:L}^{(i)}$ directamente a un vector del tamaño del horizonte. Este vector se suma a la predicción no lineal obtenida del decodificador temporal. Esta estructura garantiza matemáticamente que un modelo estrictamente lineal sea siempre un subconjunto de TiDE, permitiéndole emular el comportamiento de soluciones lineales básicas cuando sea óptimo.

Ventajas y Propiedades Computacionales: La formulación de TiDE presenta importantes ventajas empíricas y computacionales respecto a los modelos dominantes basados en mecanismos de atención:

- **Escalamiento Lineal:** Al carecer de mecanismos de auto-atención, convoluciones o estructuras recurrentes, la inferencia en el codificador de TiDE escala de forma lineal respecto a la longitud del contexto (L) y del horizonte (H), con una complejidad de $\tilde{O}(n_e h^2 + hL)$ (donde h representa el tamaño oculto). En contraste, los Transformers vanilla escalan de forma cuadrática ($\tilde{O}(L^2)$), lo que resulta prohibitivo en memoria y cómputo para contextos muy largos.
- **Soporte de Covariables y Atributos:** A diferencia de otros enfoques simplificados como DLinear (que descarta los covariables temporales debido a que degradan su rendimiento lineal), TiDE integra de forma nativa atributos estáticos y covariables

complejas pasadas y futuras gracias al diseño de sus módulos de proyección y decodificación temporal.

- **Eficiencia de Entrenamiento:** Experimentalmente, se demostró que TiDE es capaz de igualar o superar la precisión de los Transformers basados en parches (como PatchTST) en grandes conjuntos de datos (por ejemplo, en el dataset Traffic), mientras opera de 5 a 10 veces más rápido en procesos de inferencia y entrenamiento

CAPÍTULO 3:

3 DESARROLLO

3.1 Descripción las variables

3.1.1 Fuentes de datos

Las principales fuentes de información corresponden a entidades locales oficiales y bases de datos internacionales. En primer lugar, se utilizan datos del Banco Central del Ecuador, específicamente las series del ITCER e ITCBR, junto con indicadores macroeconómicos como el Índice de Precios al Consumidor (IPC) y tasas de inflación, con frecuencia mensual y que tienen cobertura desde 1995 hasta 2026, pero se utilizarán únicamente los datos del 2018 hasta 2025.

La base incluye variables clave como el valor del índice, así como sus tasas de variación mensual y anual. La estructura del dataset está organizada por año y mes, lo que permite construir una serie temporal continua y adecuada para modelos de pronóstico. También, el archivo contiene hojas complementarias con desagregaciones por regiones (América, Europa, Asia) y componentes del índice, lo cual abre la posibilidad de enriquecer el análisis mediante variables adicionales o enfoques multivariados.

Un aspecto relevante del dataset es que no solo proporciona el nivel del ITCER, sino también métricas derivadas como la tasa de variación mensual, que captura cambios de corto plazo, y la tasa de variación anual, que refleja tendencias de mediano plazo. Esto resulta especialmente útil para modelos de *Transformers temporales*, ya que permite incorporar diferentes representaciones de la dinámica de la serie (nivel y cambios), mejorando la capacidad del modelo para capturar patrones complejos y realizar pronósticos multi-horizonte.

3.1.2 Definición de variables del modelo

El modelo propuesto se estructura a partir de un conjunto de variables organizadas en tres categorías: variable objetivo, variables explicativas macroeconómicas y variables de calendario. Esta estructura permite capturar tanto la dinámica propia del ITCER como los factores externos que influyen en su comportamiento. Asimismo, se incorporan transformaciones de las variables, tales como rezagos y diferencias, con el objetivo de mejorar la capacidad predictiva del modelo y facilitar la captura de patrones temporales relevantes.

3.1.3 Variables objetivo

La variable objetivo del modelo es el Índice de Tipo de Cambio Efectivo Real (ITCER), publicado por el Banco Central del Ecuador. Este indicador refleja la competitividad internacional del país y, en el caso ecuatoriano, depende principalmente de la inflación relativa debido al régimen de dolarización. El ITCER se modela como una serie temporal mensual, y su pronóstico se realiza en múltiples horizontes de tiempo, permitiendo anticipar su evolución futura y analizar cambios en la competitividad externa.

3.1.4 Variables exógenas

Debido a que el ITCER refleja principalmente cambios en la competitividad relativa de Ecuador frente a sus principales socios comerciales, se priorizaron variables asociadas con los precios internos y externos, las condiciones monetarias y cambiarias internacionales, los ingresos provenientes del sector petrolero y las condiciones de riesgo de la economía ecuatoriana. La selección de las variables exógenas se realizó considerando su relevancia económica para la evolución del Índice de Tipo de Cambio Efectivo Real (ITCER), su disponibilidad con frecuencia mensual y la factibilidad de incorporarlas al modelo de pronóstico.

Bajo estos criterios, el conjunto de variables exógenas se delimitó a cinco indicadores: Índice de Precios al Consumidor (IPC) de Ecuador, IPC de Estados Unidos, USD NEER, precio mensual del petróleo WTI y riesgo país EMBI.

El IPC de Ecuador se incorporó para representar la evolución de los precios internos, mientras que el IPC de Estados Unidos permite aproximar los cambios en los precios externos y, en conjunto con el indicador nacional, representar parte de la dinámica de los precios relativos. El USD NEER se seleccionó para representar la evolución del dólar estadounidense frente a otras monedas. Esta variable es especialmente relevante para Ecuador debido a la dolarización de la economía, ya que el país no cuenta con un tipo de cambio nominal propio que pueda utilizarse como variable explicativa.

El precio del petróleo WTI se incluyó debido a la importancia del sector petrolero en la economía ecuatoriana y a la exposición del país a las variaciones de los precios internacionales del petróleo. De esta manera, la variable permite incorporar posibles choques externos relacionados con el sector petrolero. Por su parte, el riesgo país EMBI se utiliza como una aproximación de las condiciones de riesgo y de la percepción de los mercados sobre la economía ecuatoriana, aspectos que pueden reflejar cambios en las condiciones de financiamiento externo.

Por otro lado, no se incluyeron otras variables macroeconómicas como la balanza comercial, reservas internacionales, agregados fiscales o tasas de interés domésticas, por dos razones metodológicas principales. Primero, frente al tamaño muestral disponible, la inclusión de múltiples regresores consume grados de libertad e incrementa el riesgo de sobreajuste (overfitting), lo cual deteriora la capacidad de generalización y la precisión predictiva fuera de muestra (Box et al., 2015; Stock & Watson, 2007). Segundo, los

determinantes externos seleccionados (WTI, EMBI y USD NEER) ya sintetizan los principales canales de transmisión macroeconómica, evitando problemas de multicolinealidad y mayor capa de complejidad en el preprocesamiento de los datos.

En consecuencia, las variables seleccionadas, Tabla 1, permiten representar distintas dimensiones relacionadas con el comportamiento del ITCER: precios internos y externos, condiciones cambiarias internacionales, choques asociados al sector petrolero y condiciones de riesgo macrofinanciero. Esta delimitación busca mantener un equilibrio entre la relevancia económica de las variables, la disponibilidad de información y la viabilidad estadística y computacional del modelo de pronóstico multi-horizonte. Las variables no incluidas podrán ser consideradas en futuras investigaciones.

Tabla 1
Variables exógenas del modelo

Tipo	Variable	Justificación
Objetivo	ITCER Ecuador	Variable central de competitividad externa
Exógena 1	IPC Ecuador	Captura precios internos
Exógena 2	IPC Estados Unidos o inflación externa de referencia	Permite aproximar inflación relativa
Exógena 3	USD NEER o índice nominal externo equivalente	Captura variaciones nominales externas relevantes para Ecuador
Exógena 4	Precio del petróleo WTI	Aproxima choques externos relevantes para balanza externa e ingresos petroleros
Exógena 5	Riesgo país EMBI	Captura percepción de riesgo macrofinanciero
Calendario	Mes, trimestre, tendencia temporal	Captura estacionalidad y estructura temporal

Fuente: Autores

3.1.5 Variables calendario

Las variables de calendario se incluyen para capturar patrones estacionales y efectos temporales recurrentes en la serie. Estas variables comprenden información como el mes, 3 meses y 6 meses, así como posibles indicadores de estacionalidad. Su inclusión permite al modelo identificar regularidades en el comportamiento del ITCER asociadas a ciclos económicos o patrones recurrentes en el tiempo, mejorando así la precisión del pronóstico multi-horizonte.

CAPITULO 4:

4 Desarrollo del proyecto

4.1 Construcción y preprocesamiento de la base de datos

El desarrollo del trabajo inició con la recopilación y preparación de series macroeconómicas relacionadas con el comportamiento del Índice de Tipo de Cambio Efectivo Real (ITCER) de Ecuador la cual es la variable objetivo del estudio. Mientras que las variables explicativas evaluadas incluyeron el precio internacional del petróleo West Texas Intermediate (WTI), el Índice de Precios al Consumidor de Estados Unidos, el Índice de Tipo de Cambio Efectivo Nominal (NEER), el índice del dólar estadounidense (DXY), el riesgo país, la tasa de fondos federales de Estados Unidos, las exportaciones petroleras y el Índice de Términos de Intercambio (ITI).

Los datasets originales presentaban diferentes formatos temporales, unos eran mensuales, otros diarios. Para garantizar la correcta comparación entre variables y permitir su uso en modelos de series temporales, todas las series fueron estandarizadas a frecuencia mensual.

Para las variables con frecuencia diaria se utilizó el promedio de los valores observados durante cada mes. La fecha resultante de cada observación fue asignada al último día de cada mes.

Diseño experimental

El diseño experimental se planteó como una evaluación fuera de muestra con partición temporal. La serie del ITCER tiene una estructura secuencial y debe conservarse el orden cronológico de los datos. El periodo de entrenamiento incluyó la información disponible hasta junio de 2025, mientras que el periodo de prueba comprendió desde julio de 2025 hasta febrero de 2026. Esta división se mantuvo como criterio general para los modelos

evaluados, aunque en algunos casos el número efectivo de observaciones cambió por la construcción de rezagos o ventanas históricas.

Los modelos se compararon en los horizontes $h1$, $h3$ y $h6$, que corresponden a pronósticos a 1, 3 y 6 meses. Debido a que la base de datos llega hasta febrero de 2026, el número de observaciones reales disponibles cambia según el horizonte: siete observaciones para $h1$, cinco para $h3$ y dos para $h6$.

Para comparar los modelos, se utilizaron datos posteriores al periodo de entrenamiento y las mismas métricas siempre que la estructura del modelo lo permitió. El MAE se tomó como métrica principal porque muestra el error promedio en las mismas unidades del ITCER. El RMSE se usó como métrica complementaria, ya que penaliza más los errores grandes. Además, se reportó MAPE o sMAPE para expresar el error en términos porcentuales. El R^2 se analiza de igual forma, complementariamente, especialmente en $h6$, debido al número reducido de observaciones fuera de muestra.

La selección de hiperparámetros en los modelos se realizó con un criterio que busca reducir el riesgo de sobreajuste y precisión. En los modelos de aprendizaje profundo se usaron ventanas históricas de 12 meses, validación temporal, detención temprana y regularización. En LightGBM se evaluó una grilla acotada de tamaños de modelo y parámetros de regularización. La selección final se basó en el desempeño de validación o en configuraciones definidas antes de evaluar el conjunto de prueba, evitando ajustar los modelos directamente sobre los datos de prueba.

4.2 Modelos base para comparación

4.2.1 Estimación con modelo Naive Persistence

Como primer modelo de referencia se implementó el modelo Naive Persistence para el pronóstico del Índice de Tipo de Cambio Real de Ecuador, ITCER. Este modelo parte de

una idea o regla simple: el valor futuro de la serie se considera igual al último valor observado. En el proyecto, esta regla se aplicó para los tres horizontes de predicción definidos: un mes, tres meses y seis meses.

Este modelo no incorpora variables exógenas ni estima parámetros. Su regla de pronóstico consiste en asumir que el valor futuro del ITCER será igual al último valor observado. En el proyecto, esta regla se aplicó para tres horizontes de predicción: 1, 3 y 6 meses.

La expresión general del modelo es:

$$\widehat{y_{t+h}} = y_t$$

4.2.2 Datos utilizados y partición temporal

El modelo se evaluó sobre la base mensual consolidada del ITCER y variables macroeconómicas para el periodo enero de 2018 a febrero de 2026. Aunque Naive Persistence no requiere entrenamiento, se mantuvo la misma partición temporal definida para el resto de los modelos, con el fin de conservar una comparación bajo las mismas condiciones. Como se muestra en la Tabla 2.

Tabla 2
Elaboración propia a partir del dataset de modelado.

Partición	Periodo	Observaciones	Uso dentro del proyecto
Entrenamiento	2018-01 a 2025-06	90	Referencia temporal común para los modelos supervisados
Prueba	2025-07 a 2026-02	8	Evaluación final comparable entre modelos

Fuente: Autores

Para cada fecha del periodo de prueba se tomó el ITCER observado en el mes de origen y se lo uso como pronóstico para los horizontes de 1, 3 y 6 meses. La Tabla 3 resume

los valores reales y las predicciones generadas por el modelo. En las últimas fechas de origen algunos valores reales futuros no se encuentran disponibles, porque la base llega hasta febrero de 2026 , por ello, las métricas se calculan con las observaciones disponibles en cada horizonte.

Tabla 3
Métricas de modelo Naive por cada horizonte

Fecha de origen	ITCER observado	Real h1	Pred h1	Real h3	Pred h3	Real h6	Pred h6
2025-07	105,04	104,65	105,0399	105,29	105,04	107,38	105,04
2025-08	104,65	105,31	104,6479	105,70	104,65	108,1319	104,65
2025-09	105,31	105,29	105,306	106,60	105,31	-	105,31
2025-10	105,29	105,70	105,2925	107,38	105,29	-	105,29
2025-11	105,70	106,60	105,7001	108,13	105,70	-	105,70
2025-12	106,60	107,38	106,5963	-	106,60	-	106,60
2026-01	107,38	108,13	107,38	-	107,38	-	107,38
2026-02	108,13	-	108,1319	-	108,13	-	108,13

Fuente: Autores

Las métricas de evaluación muestran que el error aumenta conforme se amplía el horizonte de predicción. Esto es esperable en un modelo de persistencia, ya que la regla de repetir el último valor observado pierde precisión cuando se intenta anticipar periodos más alejados. Debido a la disponibilidad de valores futuros hasta febrero de 2026, cada horizonte se evalúa con un numero distinto de observaciones. Los resultados se muestran en Tabla 4.

Tabla 4
Métricas de modelos Naive Persistence

Modelo	Horizonte	Observaciones	MAE	RMSE	sMAPE
Naive Persistence	h1	7	0,56	0,6252	0,01
Naive Persistence	h3	5	1,42	1,6191	0,01

Naive Persistence	h6	2	2,91	2,9676	0,03
-------------------	----	---	------	--------	------

Fuente: Autores

El error absoluto promedio paso de 0,5576 en el horizonte de un mes a 2,9120 en el horizonte de seis meses. El mayor error individual se observó en h6, con un valor de 3,4840. Esto se relaciona con el comportamiento creciente del ITCER hacia el cierre del periodo de prueba, donde el modelo tendió a subestimar los valores futuros como se detalla en Tabla 5.

Tabla 5

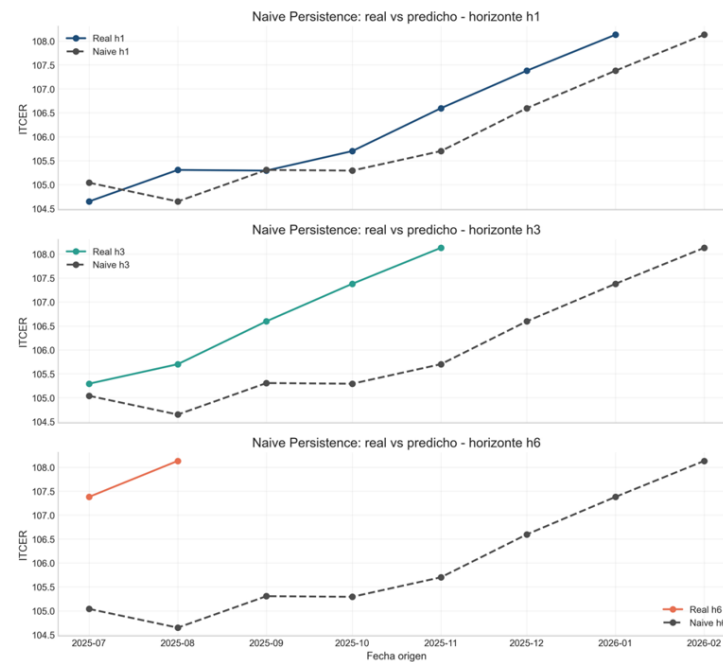
Errores obtenidos en el modelo Naive

Estadístico	Error abs. h1	Error abs. h3	Error abs. h6
Media	0,56	1,4229	2,912
Desviación estándar	0,31	0,8638	0,8089
Mínimo	0,01	0,2526	2,34
Mediana	0,658	1,2903	2,912
Máximo	0,90	2,4318	3,484

Fuente: Autores

En la Figura 3, se observa la comparación consolidada entre los diferentes pronósticos. Esta comparación confirma que el error aumenta conforme se amplía el horizonte de predicción. En h1, el modelo conserva una cercanía mayor con los valores reales; en h3 la brecha se incrementa; y en h6 la limitación del enfoque se vuelve más visible, porque la persistencia no incorpora tendencia ni información adicional para ajustar el pronóstico.

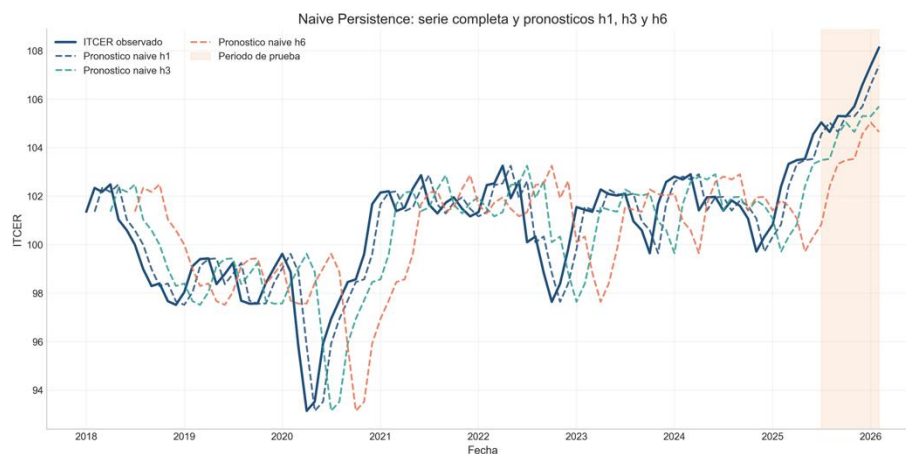
Figura 3
Comparación de predicciones para modelos Naive



Fuente: Autores

La Figura 4, muestra la serie completa y los pronósticos. La grafica nos permite ubicar el periodo de prueba dentro de la trayectoria histórica del ITCER. Las líneas punteadas muestran los pronósticos generados por la regla de persistencia para los horizontes h1, h3 y h6. Al repetir el ultimo valor observado, el modelo reproduce la dinámica reciente con rezago y no anticipa plenamente el incremento observado hacia el cierre de la muestra.

Figura 4
Pronósticos del modelo de persistencia para los horizontes h1, h3 y h6



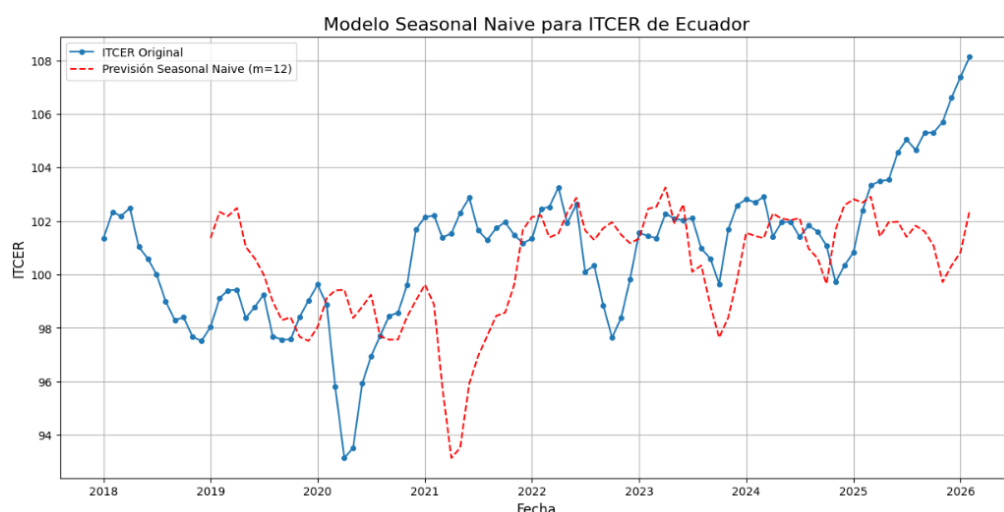
4.2.3 Estimación con modelo Seasonal Naive

El modelo Seasonal Naive (Ingenuo Estacional) constituye una de las líneas base más utilizadas en problemas de pronóstico de series temporales con comportamiento estacional. Su principio consiste en asumir que el valor futuro será igual al registrado durante el mismo período del ciclo estacional anterior.

En esta investigación, debido a que el ITCER corresponde a una serie mensual con una estacionalidad de doce meses, el modelo estima que el valor de un determinado mes será igual al observado en ese mismo mes del año anterior. Por ejemplo, el valor pronosticado para enero de 2026 corresponde al valor registrado en enero de 2025.

Figura 5

Pronóstico del modelo Seasonal Naive para el ITCER



Fuente: Autores

La Figura 5 muestra el ajuste obtenido por el modelo sobre la serie histórica del ITCER. Se observa que el modelo reproduce adecuadamente el patrón estacional utilizando la información del año anterior; sin embargo, presenta limitaciones para representar cambios bruscos, tendencias emergentes y variaciones estructurales propias de la dinámica económica. Como consecuencia, los errores de predicción aumentan durante los períodos de mayor volatilidad.

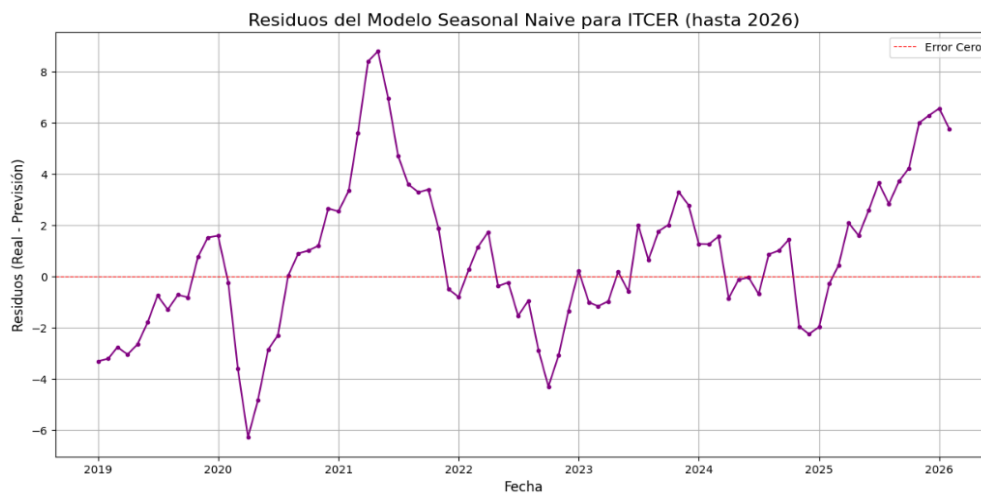
4.2.3.1 Evaluación del modelo

Como primera etapa, el modelo fue evaluado utilizando la serie original del ITCER, obteniéndose las siguientes métricas de desempeño:

- Error Cuadrático Medio (RMSE): 13.43
- Error Absoluto Medio (MAE): 11.36

Estas métricas constituyen la línea base de comparación para evaluar posteriormente el desempeño de los modelos LSTM Multi-horizonte y Temporal Fusion Transformer

Figura 6
Residuos del modelo Seasonal Naive



Fuente: Autores

El análisis de los residuos mostrada en Figura 6 evidencia que los mayores errores se concentran en períodos donde la serie presenta variaciones abruptas. Este comportamiento sugiere que el modelo responde adecuadamente al patrón estacional predominante, pero posee una capacidad limitada para representar cambios inesperados en el comportamiento del ITCER.

4.2.3.2 Influencia de los valores atípicos

Con el fin de evaluar el efecto de los valores atípicos sobre el desempeño del modelo, se repitió el proceso de entrenamiento utilizando la serie previamente depurada mediante el método del rango intercuartílico (IQR).

Los resultados obtenidos Tabla 6 muestran una reducción del error de predicción respecto a la serie original, lo que indica que los valores extremos ejercían una influencia importante sobre el ajuste del modelo. En consecuencia, la eliminación de estas observaciones permitió obtener una representación más estable del comportamiento predominante del ITCER.

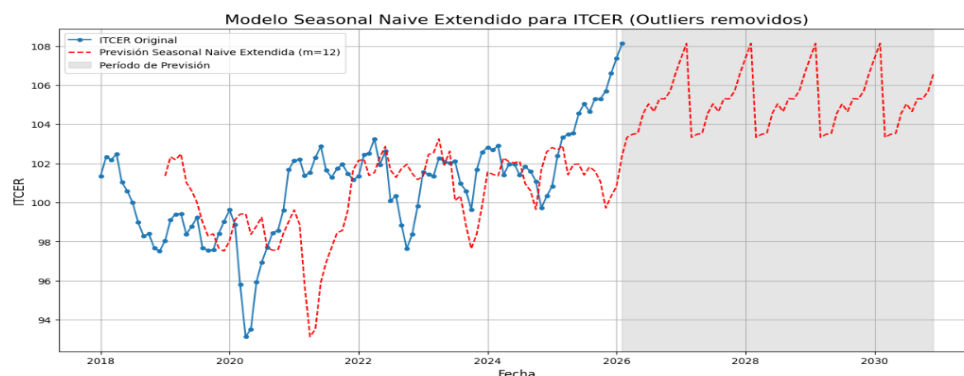
Tabla 6
Resultados métricas con y sin outliers

Métrica	Con outliers	Sin outliers
Error Cuadrático Medio (MSE)	9.12	6.58
Raíz del Error Cuadrático Medio (RMSE)	3.02	2.57
Error Absoluto Medio (MAE)	2.32	2.05

Fuente: Autores

4.2.3.3 Comportamiento del modelo en un horizonte de largo plazo

Con el propósito de analizar el comportamiento del modelo en un horizonte de predicción más amplio, el pronóstico fue extendido hasta el año 2030.

Figura 7*Pronóstico del modelo Seasonal Naive extendido hasta el año 2030*

Fuentes: Autores

La proyección muestra que el modelo mantiene de forma repetitiva el patrón estacional observado en los datos históricos Figura 7, debido a que sus predicciones se basan exclusivamente en los valores registrados durante el mismo período del año anterior. A medida que aumenta el horizonte de pronóstico, esta estrategia limita la capacidad del modelo para adaptarse a cambios estructurales o variaciones en las condiciones macroeconómicas, reduciendo progresivamente la precisión de las estimaciones.

En conjunto, los resultados obtenidos confirman que el modelo Seasonal Naive constituye una referencia adecuada como línea base de comparación. No obstante, sus limitaciones para capturar relaciones no lineales y adaptarse a cambios en la dinámica de la serie justifican la utilización de modelos de aprendizaje profundo, como la arquitectura LSTM Multi-horizonte y el Temporal Fusion Transformer, desarrollados en las siguientes secciones.

4.2.4 Estimación modelo ETS

4.2.4.1 Implementación de modelos clásicos de suavizamiento exponencial

Estos modelos fueron seleccionados debido a su capacidad para representar patrones de nivel, tendencia y estacionalidad, sin requerir variables explicativas externas. Como

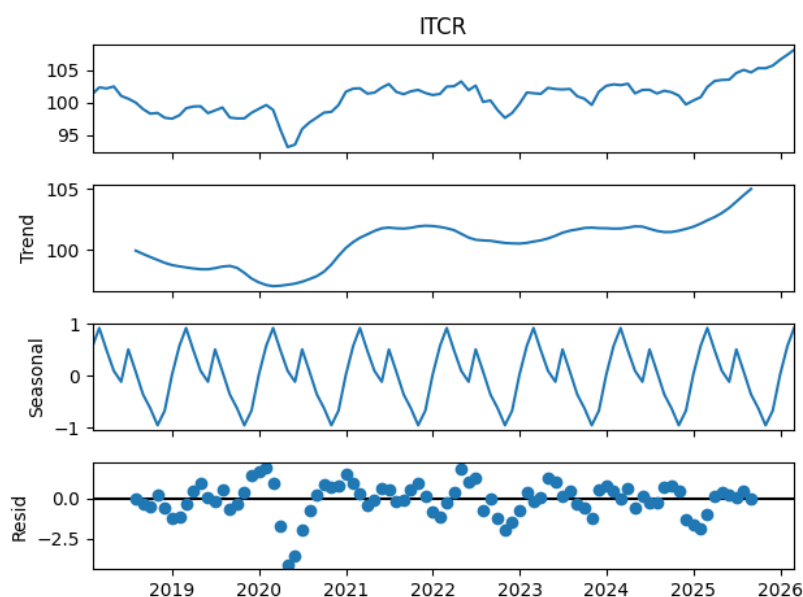
modelo de referencia se implementó un modelo de Suavizamiento Exponencial Triple (Holt-Winters o ETS) con el propósito de modelar la evolución temporal del ITCR utilizando únicamente la información histórica de la propia serie.

El modelo se configuró considerando: tendencia aditiva, componente estacional aditiva, período estacional de 12 meses, tendencia amortiguada (damped trend).

Posteriormente, la serie fue dividida en un conjunto de entrenamiento (80 %) y uno de validación correspondiente a los últimos doce meses, con el fin de evaluar la capacidad predictiva del modelo mediante datos no utilizados durante el entrenamiento.

Figura 8

Descomposición de ITCR: Tendencia, Estacionalidad y Residuos



Fuente: Autores

La Figura 8 corresponde a la descomposición de la serie temporal del ITCR, donde se separa la serie original en sus componentes de tendencia, estacionalidad y residuales. Este análisis permite comprender mejor el comportamiento del índice y evaluar si el modelo ETS es adecuado para representar su dinámica.

La descomposición evidencia que el comportamiento del ITCR está dominado principalmente por la tendencia de largo plazo, mientras que la estacionalidad tiene una influencia reducida y estable. Además, los residuos muestran una distribución aleatoria alrededor de cero, lo que indica que la mayor parte de la información de la serie ha sido capturada por los componentes sistemáticos del modelo.

4.2.4.2 Entrenamiento del modelo ETS

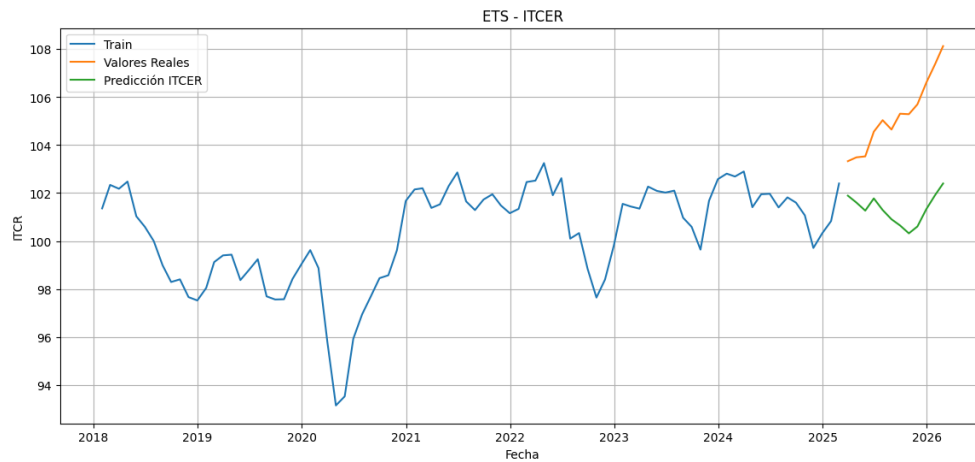
El modelo se especificó mediante una estructura de tendencia y estacionalidad aditivas, con un período estacional de 12 meses. Además, se incorporó una tendencia amortiguada (damped trend), de manera que el crecimiento o decrecimiento estimado se reduzca gradualmente conforme aumenta el horizonte de pronóstico. Los parámetros de suavizamiento del nivel, la tendencia y la estacionalidad se estimaron automáticamente durante el ajuste del modelo mediante el procedimiento de optimización implementado en statsmodels.

La especificación utilizada puede representarse como:

$$ETS(A, A_d, A)_{\{12\}}$$

Donde el primer componente representa la estructura aditiva del error, el segundo corresponde a una tendencia aditiva amortiguada y el tercero a una componente estacional aditiva con periodicidad de 12 meses. Una vez entrenado el modelo, se generaron pronósticos para los 12 meses del conjunto de prueba y se evaluó su desempeño mediante MAE, RMSE, MAPE. La figura 9 muestra el pronóstico del modelo ETS del ITCER.

Figura 9
Evaluación del modelo ETS para ITCER

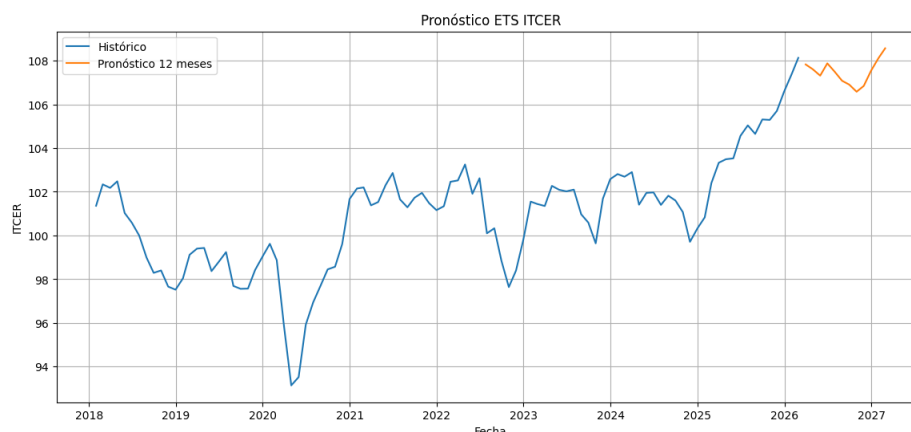


Fuente: Autores

4.2.4.3 Evaluación del modelo ETS

Los resultados muestran que el modelo ETS ofrece un buen desempeño global, alcanzando un MAPE de 3.71 %, un MAE de 3.92 y un RMSE de 4.18, lo que evidencia una alta precisión en términos generales. La cercanía entre el MAE y el RMSE indica que los errores son consistentes y no están dominados por observaciones atípicas.

No obstante, el análisis gráfico en Figura 9 revela que el modelo subestima el incremento observado en el período de prueba, manteniendo un pronóstico cercano al promedio histórico mientras que los valores reales presentan una tendencia creciente.

Figura 10*Pronóstico de ITCER aplicando ETS*

Fuente: Autores

4.2.4.4 Comportamiento del modelo ETS en un horizonte de largo plazo

En la Figura 10 se presenta el pronóstico del modelo ETS para los próximos 12 meses, el cual muestra un comportamiento relativamente estable del ITCR, con una ligera disminución en los primeros meses y una recuperación gradual hacia el final del horizonte de predicción.

Sin embargo, debido a la tendencia del modelo a suavizar las variaciones recientes, los resultados deben interpretarse como una referencia basada en patrones históricos y complementarse con el análisis del contexto macroeconómico y la comparación con otros modelos de pronóstico evaluados en la investigación.

4.2.5 Estimación modelo SARIMAX

4.2.5.1 Análisis de estacionariedad

Antes de estimar los modelos SARIMAX se evaluó la estacionariedad de las series mediante la prueba aumentada de Dickey-Fuller (ADF). Esta prueba permitió identificar si las variables presentaban una raíz unitaria y, por tanto, si requerían transformaciones antes de ser incorporadas al modelo. Los resultados se muestran en la Tabla 7.

Tabla 7*Resultados de la prueba aumentada de Dickey-Fuller (ADF) para ITCER*

ITCR	ADF	p-value
Serie original	0.4711	0.983968
Primera diferencia	-5.1241	0.000013

Fuente: Autores

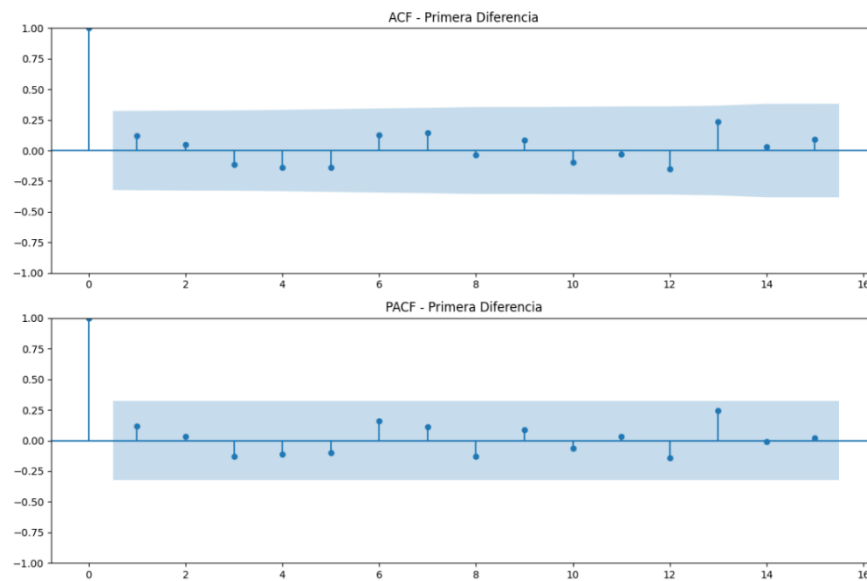
La prueba de Dickey-Fuller Aumentada aplicada al ITCR en niveles obtuvo un estadístico de (0.4711) y un valor ($p=0.983968$), por lo que no se rechazó la hipótesis nula de presencia de raíz unitaria. En consecuencia, la serie del ITCR en niveles no presentó estacionariedad. Posteriormente, se aplicó una primera diferencia a la serie, obteniéndose un estadístico ADF de (-5.1241) y un valor ($p<0.001$). Este resultado permitió rechazar la hipótesis nula de raíz unitaria y confirmar que la primera diferencia del ITCR es estacionaria.

4.2.4.2 Identificación y selección del modelo ARIMA

Una vez obtenida una serie estacionaria, se analizaron las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), las cuales permitieron identificar el posible orden de los componentes autorregresivos y de medias móviles, cuyos resultados se observan en la Figura 11.

Posteriormente se evaluaron múltiples combinaciones de parámetros (p,d,q), seleccionando el modelo con menor criterio de información de Akaike (AIC).

Figura 11
Análisis ACF y PACF de ITCER



Fuente: Autores

El análisis ACF/PACF en la Figura 11, sugirió modelos con bajo orden, principalmente estructuras AR(1) y MA de bajo orden. Posteriormente se evaluaron múltiples combinaciones mediante criterios de información y desempeño predictivo, obteniendo como mejor especificación SARIMA(0,1,2). Los resultados de los experimentos mencionados se muestran en la Tabla 8 y su representación gráfica en la Figura 12.

Tabla 8
Métricas de experimentos SARIMA

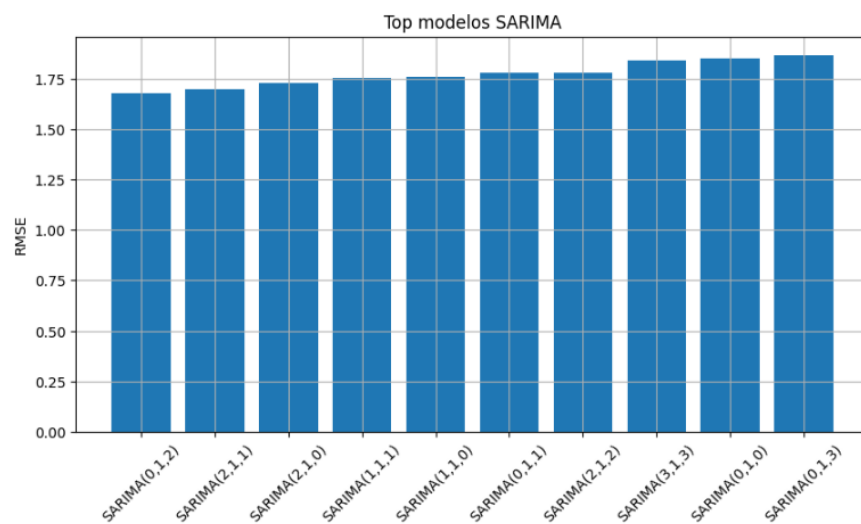
Modelo	AIC	BIC	MAE	RMSE
SARIMA(0,1,2)	69.280172	73.054462	1.265333	1.676581
SARIMA(2,1,1)	73.862464	79.045812	1.291149	1.701416
SARIMA(2,1,0)	72.160923	76.048433	1.314263	1.730931
SARIMA(1,1,1)	72.184703	76.072213	1.335124	1.754723

SARIMA(1,1,0)	71.632225	74.296634	1.340299	1.760521
SARIMA(0,1,1)	70.220176	72.811850	1.358664	1.780250
SARIMA(2,1,2)	69.330494	75.620977	1.349939	1.781171
SARIMA(3,1,3)	70.261829	78.793960	1.417508	1.840642
SARIMA(0,1,0)	69.928061	71.260265	1.452500	1.853793
SARIMA(0,1,3)	69.323561	74.199064	1.445859	1.866432
SARIMA(3,1,0)	70.485237	75.517623	1.484006	1.895532
SARIMA(3,1,1)	72.459654	78.750137	1.552172	1.966542
SARIMA(1,1,2)	71.575823	76.608209	2.382848	2.781746
SARIMA(1,1,3)	70.755519	76.849898	3.040375	3.484955
SARIMA(3,1,2)	74.409193	81.957772	3.578817	3.972314
SARIMA(2,1,3)	71.784460	79.097715	3.618460	3.993154

Fuente: Autores

Figura 12

Modelos SARIMA ordenados por RMSE



Fuente: Autores

Se utilizó una diferenciación de orden uno en la estimación de los modelos SARIMAX. Por esta razón, los modelos SARIMAX fueron estimados utilizando la variación mensual del ITCR como variable objetivo.

También se evaluó la posibilidad de que ciertas variables macroeconómicas afecten al ITCER con retraso. La inclusión de rezagos permite capturar efectos que no se reflejan de manera inmediata sobre el ITCER, especialmente en variables cuyo impacto puede transmitirse a través de canales fiscales, monetarios o comerciales.

La evaluación de los modelos se realizó mediante una división temporal de la muestra: el 80 % de las observaciones se destinó al entrenamiento y el 20 % restante a la prueba. Esta partición respeta el orden cronológico de la serie y permite evaluar la capacidad predictiva del modelo sobre períodos no utilizados durante su estimación.

El desempeño de los modelos fue evaluado mediante el Error Absoluto Medio (MAE), la Raíz del Error Cuadrático Medio (RMSE), el Error Porcentual Absoluto Medio (MAPE) y R^2 . Adicionalmente, se generaron gráficos comparativos entre los valores reales y las predicciones.

4.2.4.3 Implementación de modelos SARIMAX con variables macroeconómicas

Experimento 1

Inicialmente se incorporaron diversas variables macroeconómicas al modelo SARIMAX. Posteriormente se evaluó la significancia estadística de cada una mediante sus coeficientes estimados y valores p, estos resultados se detallan en la Tabla 9.

Tabla 9
Coeficientes de variables macroeconómicas

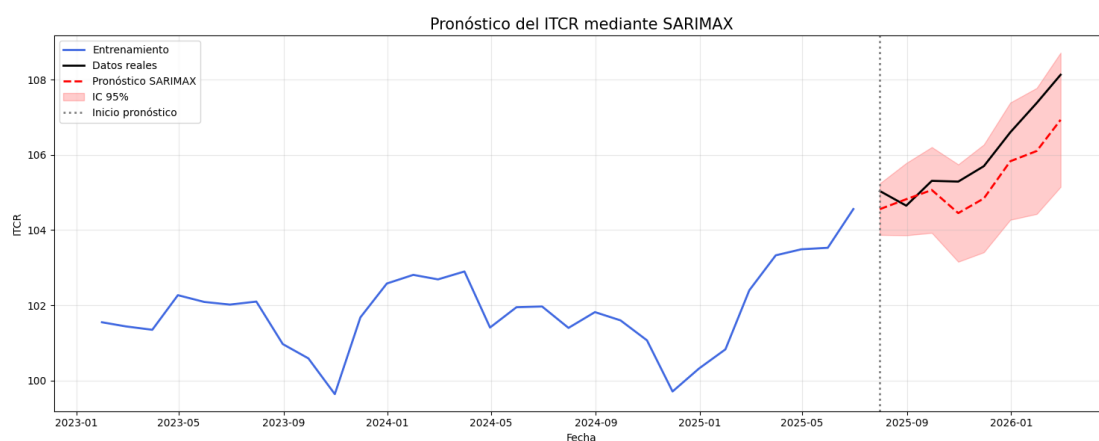
Variable	Coeficiente	Valor p
IPCE	-0.88480	0.00019
NEER	-0.49597	0.00002

IPC	0.35485	0.01929
FED	0.21329	0.83898
sigma2	0.12192	0.09869
ma.L2	-0.08847	0.80458
ITI	-0.05639	0.29976
WTI	0.02760	0.43752
ma.L1	-0.01492	0.96598
RIESGO	-0.00028	0.75814
EXPORTA	0.00000	0.47415

Fuente: Autores

Figura 13

Resultados entrenamiento SARIMAX con todas las variables macroeconómicas



Fuente: Autores

La Figura 13 muestra que el modelo SARIMAX logra seguir de manera adecuada la tendencia general del ITCR y proyecta un crecimiento hacia el final del período analizado. Sin embargo, se observan algunas diferencias respecto a los valores reales, especialmente ante cambios de corto plazo. Además, el intervalo de confianza se amplía a medida que aumenta el horizonte de pronóstico, reflejando una mayor incertidumbre en las estimaciones. En general, el modelo representa razonablemente el comportamiento del ITCR, aunque su precisión disminuye en períodos más alejados.

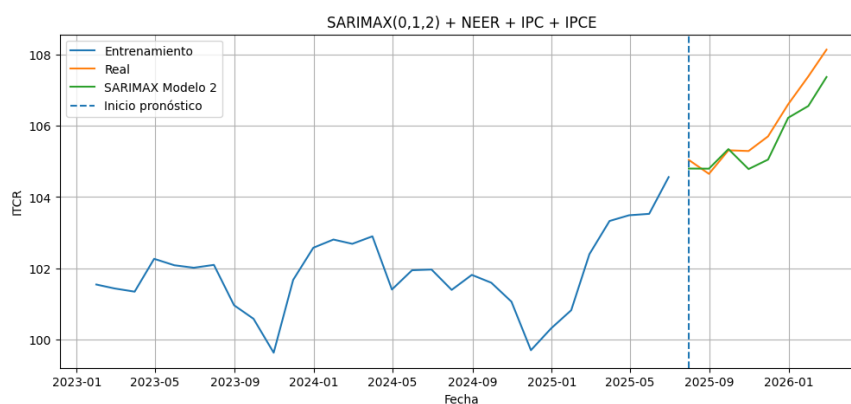
Tabla 10
Métricas SARIMAX experimento 1

Métrica	Valor
MAE	0.730
RMSE	0.823
MAPE	0.69%
R ²	0.4897

Fuente: Autores

En la Tabla 10 se muestran las métricas obtenidas, Como resultado del análisis se identificó que las variables con mayor aporte al comportamiento del ITCR fueron: NEER, IPC, IPCE, ITI (con rezago). Las variables que no aportaban información significativa fueron descartadas para reducir la complejidad del modelo.

Figura 14
Resultados entrenamiento SARIMAX con variables más representativas



Fuente: Autores

En la Figura 14 se aprecia que, durante el período de entrenamiento, el ITCR muestra algunas fluctuaciones, con una recuperación hacia mediados de 2025. En el período de pronóstico, el modelo sigue de manera cercana el comportamiento observado y reproduce la tendencia creciente del índice. Aunque existen pequeñas diferencias entre los valores reales y estimados, la trayectoria general es similar. Esto sugiere que la selección de las variables más

representativas permite al modelo captar adecuadamente la dinámica reciente del ITCR y generar pronósticos consistentes con los valores observados. Los resultados obtenidos en este experimento se presentan en la Tabla 11.

Tabla 11
Métricas SARIMAX con variables más representativas

Métrica	Valor
MAE	0.4440
RMSE	0.5218
MAPE	0.42%
R ²	0.7948

Fuente: Autores

4.2.4.4 Análisis y selección de rezagos de las variables exógenas

Experimento 2

Con el objetivo de capturar efectos retardados de las variables macroeconómicas, se construyeron rezagos de uno, dos y tres períodos.

Posteriormente se calculó la correlación entre cada rezago y la variación del ITCR, observándose que el rezago de dos meses del ITI presentaba la mayor relación con la variable objetivo. Por esta razón, el modelo final incorporó la variable ITI_lag2.

Se evaluó la incorporación de rezagos temporales en las variables exógenas con el objetivo de capturar efectos económicos que no se producen inmediatamente. Se analizaron rezagos entre uno y tres meses para las variables macroeconómicas y se seleccionaron aquellas con mayor relación con la variación del ITCR. Los coeficientes se presentan en la Tabla 12.

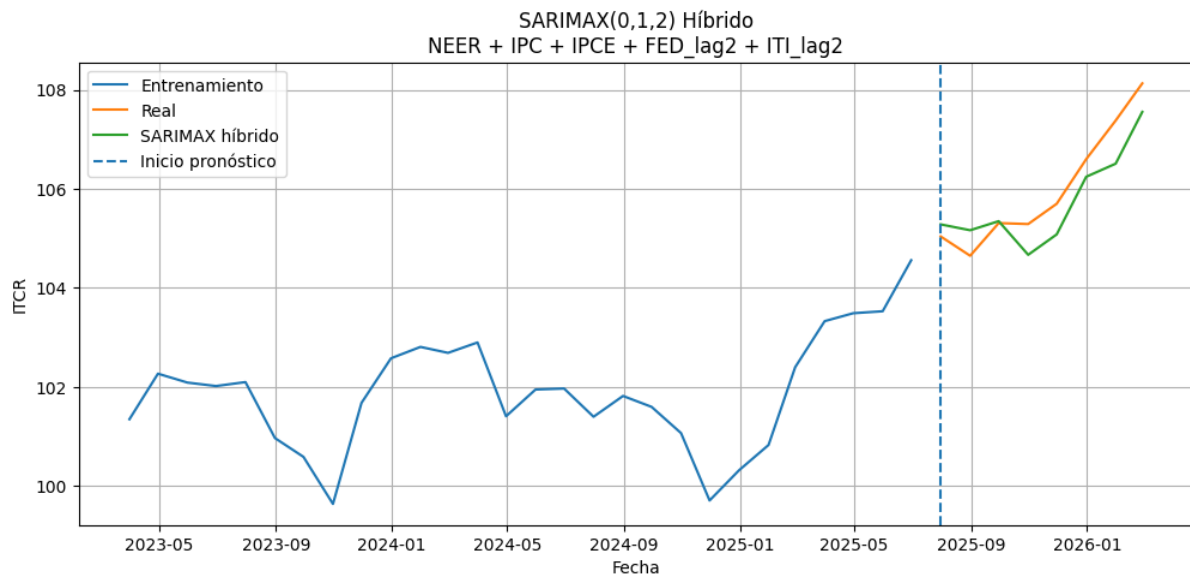
Tabla 12
Coefficientes de rezagos de variables macroeconómicas

Variable	Valor	Variable	Valor
----------	-------	----------	-------

FED_lag2	-0.387197	IPCE_lag3	0.220814	
FED_lag1		-0.380198	RIESGO_lag3	0.216301
FED_lag3		-0.350622	ITI_lag1	0.204168
NEER_lag2		0.342383	RIESGO_lag1	0.200595
NEER_lag1		0.337444	RIESGO_lag2	0.154808
ITI_lag2		0.333137	WTI_lag3	-0.136039
ITI_lag3		0.289754	IPCE_lag2	0.095435
IPC_lag2		0.270273	WTI_lag2	-0.080223
IPC_lag3		0.265591	IPCE_lag1	0.079970
IPC_lag1		0.261921	EXPORTA_lag2	-0.023575
WTI_lag1		-0.232857	EXPORTA_lag1	-0.019158
NEER_lag3	0.221497	EXPORTA_lag3	-0.003704	

Fuente: Autores

Inicialmente se identificó un modelo SARIMA(0,1,2) mediante comparación de criterios de información y error predictivo. Posteriormente se incorporaron variables macroeconómicas contemporáneas identificadas mediante análisis estadístico. Finalmente, mediante análisis de correlaciones cruzadas se evaluaron rezagos económicos, incorporándose únicamente ITI_lag2 debido a su significancia estadística y mejora en capacidad predictiva, evitando sobre parametrización debido al tamaño reducido de la muestra.

Figura 15*Resultados Entrenamiento SARIMAX híbrido*

Fuente: Autores

La Figura 15 presenta la evolución del ITCER durante el periodo de entrenamiento y el periodo de pronóstico del modelo SARIMAX(0,1,2) híbrido. Durante el periodo histórico se observa un comportamiento fluctuante, con algunos descensos y posteriores recuperaciones. A partir del inicio del pronóstico, las estimaciones del modelo mantienen una trayectoria similar a los valores reales, reproduciendo adecuadamente las principales variaciones de la serie.

Además, hacia el final del periodo se observa una tendencia creciente tanto en los valores reales como en los pronosticados, aunque con pequeñas diferencias entre ambas series.

Se evidencia una adecuada capacidad del modelo para seguir la dinámica y tendencia del ITCER, las métricas obtenidas en este experimento se muestran en la Tabla 13.

Tabla 13*Métricas SARIMAX híbrido*

Métrica	Valor
MAE	0.4798
RMSE	0.5374
MAPE	0.45%

$$\frac{R^2}{0.7823}$$

Fuente: Autores

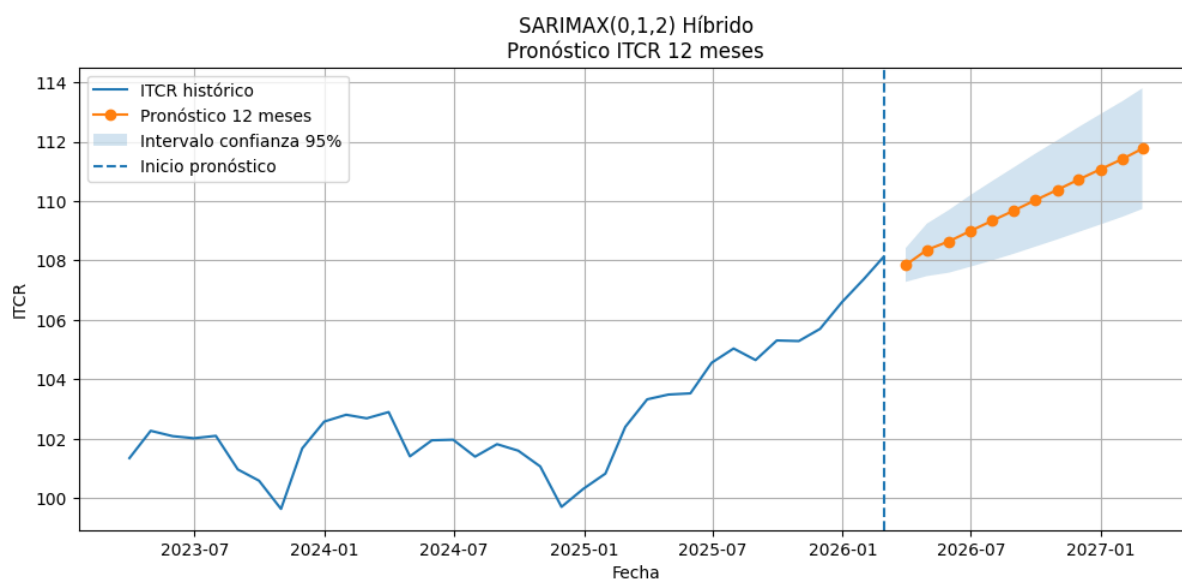
4.2.4.5 Selección del modelo SARIMAX final

Después de evaluar múltiples configuraciones mediante el criterio de información de Akaike (AIC), el análisis de significancia estadística de las variables exógenas y las métricas de validación (MAE, RMSE, MAPE y R^2), se seleccionó un modelo SARIMAX(0,1,2) con las variables exógenas NEER, IPC, IPCE e ITI con un rezago de dos periodos. Esta configuración logró un equilibrio entre precisión predictiva y simplicidad estructural, incorporando tanto la dependencia temporal del ITCR como la influencia de variables macroeconómicas relevantes. La inclusión del rezago del ITI permitió representar efectos económicos diferidos que no son capturados por las variables contemporáneas, mejorando la capacidad explicativa del modelo sin incrementar excesivamente su complejidad.

4.2.4.6 Comportamiento del modelo SARIMAX en un horizonte de largo plazo

La Figura 16 presenta el pronóstico del ITCER para los 12 meses siguientes mediante el modelo SARIMAX(0,1,2) híbrido. Se observa una tendencia creciente en los valores proyectados, pasando aproximadamente de 107,8 a 111,8, lo que indica una expectativa de incremento gradual del ITCR durante el horizonte de pronóstico.

El intervalo de confianza del 95 % se amplía progresivamente a medida que aumenta el horizonte temporal, reflejando una mayor incertidumbre asociada a las proyecciones de largo plazo. En conclusión, el modelo proyecta un comportamiento ascendente del ITCER, acompañado de un incremento gradual de la incertidumbre en las estimaciones.

Figura 16*Pronóstico SARIMAX híbrido 12 meses para ITCER*

Fuente: Autores

4.2.6 LightGBM directo multi-horizonte

El modelo LightGBM directo multi-horizonte se utilizó para evaluar un enfoque de aprendizaje automático en el pronóstico del ITCER. A diferencia de un modelo univariante, este modelo permite incorporar rezagos de la serie y variables macroeconómicas para estimar el comportamiento futuro del índice.

El modelo no predice directamente el nivel futuro del ITCER. En su lugar, estima el cambio esperado entre el valor observado en la fecha de origen y el valor correspondiente a cada horizonte de predicción. Luego, el pronóstico final se obtiene sumando ese cambio estimado al ITCER observado en el mes de origen.

Se entrenaron modelos independientes para los horizontes de 1, 3 y 6 meses. También se compararon dos conjuntos de variables: uno compuesto por el ITCER y sus rezagos, y otro que incorpora variables macroeconómicas y de calendario. Esta comparación permite revisar si la información adicional mejora el desempeño del modelo fuera de muestra. Para este

modelo, se utilizó la diferencia entre el valor actual de la serie y su valor rezagado, ya que se obtuvieron mejores resultados que al trabajar con la serie original.

La transformación del objetivo fue:

$$target_{t,h} = y_{t+h} - y_t$$

El pronóstico final del nivel del ITCER se reconstruye como:

$$\widehat{y_{t+h}} = y_t + \widehat{target_{t,h}}$$

Donde,

- y_t ITCER observado en la fecha de origen
- y_{t+h} ITCER observado en el horizonte futuro
- $target_{t,h}$ Cambio futuro del ITCER respecto al valor actual
- $\widehat{y_{t+h}}$ Pronostico final del nivel del ITCER
- h Horizonte de predicción: 1, 3 o 6 meses

4.2.6.1 Datos utilizados y partición temporal

De la misma manera que los modelos anteriores, el modelo se estimó con la base mensual de modelado del ITCER y variables macroeconómicas, los periodos empleados se detallan en la Tabla 14.

Tabla 14

Segmentación de dataset para evaluación del modelo LightGBM

Partición	Periodo	Uso dentro del proyecto
Entrenamiento	2018-01 a 2025-06	Estimación de los modelos LightGBM
Prueba	2025-07 a 2026-02	Evaluación fuera de muestra

Fuente: Autores

Pero, la fecha de inicio real del entrenamiento fue enero de 2019 porque el modelo utiliza rezagos, incluido 'ITCER_lag12'. Por esa razón, las primeras observaciones de 2018 no tienen todas las variables requeridas.

4.2.6.2 Variables utilizadas

Para este modelo se evaluaron dos conjuntos de variables. El primero incluye solo información histórica del ITCER. El segundo incorpora variables macroeconómicas y variables de calendario.

- Solo ITCER y rezagos: ITCER, ITCER_lag1, ITCER_lag3, ITCER_lag6, ITCER_lag12.
- ITCER + exogenas: Variables del conjunto anterior, y las variables IPC_ECU, IPC_USA, USD_NEER, WTI, EMBI, anio, mes, trimestre, tendencia, mes_sin, mes_cos y rezagos de variables macroeconómicas .

Donde:

- Tendencia: Variable de tiempo que representa el orden cronológico de las observaciones. Permite capturar cambios graduales asociados al paso del tiempo.
- mes_sin: Transformacion seno del mes del año. Permite representar el componente estacional mensual de forma ciclica.
- mes_cos: Transformacion coseno del mes del año. Complementa a mes_sin para capturar la estacionalidad mensual sin imponer cortes entre diciembre y enero.

4.2.6.3 Configuración del modelo

Se utilizaron configuraciones conservadoras del algoritmo LightGBM, se muestran en la Tabla 15, implementado mediante la clase LGBMRegressor de la librería lightgbm de Python, con el fin de reducir el riesgo de sobreajuste en una muestra mensual pequeña. Los modelos se definieron con `boosting_type="gbdt"` y `objective="regression_l1"`.

Tabla 15

Hiperparámetros utilizados para la estimación del modelo LightGBM

Horizonte	Configuración base	Profundidad	Hojas	Min. obs. por hoja	Regularización
-----------	--------------------	-------------	-------	--------------------	----------------

h1	250 árboles, learning rate 0,03	3	7	8	reg_alpha 0,20, reg_lambda 1,00
h3	150 árboles, learning rate 0,03	2	3	15	reg_alpha 0,50, reg_lambda 2,00
h6	30 árboles, learning rate 0,03	2	3	15	reg_alpha 0,50, reg_lambda 2,00

Fuente: Autores

Por otro lado, para la selección de hiperparámetros de LightGBM se aplicó una búsqueda controlada sobre un conjunto reducido de configuraciones, considerando el tamaño limitado de la serie mensual. Se evaluaron combinaciones asociadas al número de árboles (n_estimators), tasa de aprendizaje (learning_rate), profundidad máxima (max_depth), número de hojas (num_leaves), mínimo de observaciones por hoja (min_child_samples) y parámetros de regularización L1 y L2 (reg_alpha y reg_lambda). La selección se realizó de forma independiente para cada horizonte de pronóstico, priorizando la configuración con menor MAE en el conjunto de prueba temporal. Este procedimiento permitió ajustar la complejidad del modelo según el horizonte evaluado y reducir el riesgo de sobreajuste, evitando así una búsqueda exhaustiva para nuestra mensual con cantidad de datos modesta.

4.2.6.4 Comparación por número de arboles

La Tabla 16 muestra el MAE de prueba para cada horizonte, conjunto de variables y número de árboles. En cada fila se resalta el menor MAE.

Tabla 16
Métrica MAE de pruebas por horizonte

Horizonte	Variables	25 arboles	100 arboles	250 arboles	400 arboles
h1	Solo ITCER y rezagos	0,514	0,541	0,592	0,619
h1	ITCER + exógenas	0,454	0,492	0,47	0,484
h3	Solo ITCER y rezagos	1,661	1,994	2,164	2,265

h3	ITCER + exógenas	1,566	1,793	1,838	1,81
h6	Solo ITCER y rezagos	2,982	3,616	3,766	3,811
h6	ITCER + exógenas	2,493	2,72	2,477	2,492

Fuente: Autores

4.2.6.5 Mejores configuraciones por horizonte^[OB]

Como se observa en la Tabla 17, para h1 y h3, el menor MAE se obtuvo con variables exógenas y 25 árboles. Para h6, el mejor resultado se obtuvo con variables exógenas y 250 árboles.

Tabla 17
Mejores configuraciones por cada horizonte

Horizonte	Variables	Arboles	Observaciones	MAE	RMSE	sMAPE	
h1	ITCER exogenas	+	25	7	0,4544	0,4881	0,004281
h3	ITCER exogenas	+	25	5	1,5658	1,7197	0,014753
h6	ITCER exogenas	+	250	2	2,477	2,5274	0,023249

Fuente: Autores

La cantidad de observaciones cambia por horizonte porque la base llega hasta febrero de 2026. En h1 existen siete valores reales futuros disponibles, en h3 existen cinco y en h6 existen dos.

4.2.6.6 Predicciones

La Tabla 18 resume las predicciones usadas para las mejores configuraciones por horizonte.

Tabla 18
Resumen de predicciones realizadas

Fecha	Horizonte	ITCER	ITCER + h	Predicción	Error absoluto
2025-07	h1	105,0399	104,6479	105,2342	0,5863
2025-08	h1	104,6479	105,306	104,83	0,4759
2025-09	h1	105,306	105,2925	105,4881	0,1956
2025-10	h1	105,2925	105,7001	105,5239	0,1762
2025-11	h1	105,7001	106,5963	105,9315	0,6648
2025-12	h1	106,5963	107,38	106,8277	0,5523
2026-01	h1	107,38	108,1319	107,6018	0,53
2025-07	h3	105,0399	105,2925	104,8214	0,4711
2025-08	h3	104,6479	105,7001	104,4653	1,2348
2025-09	h3	105,306	106,5963	105,1233	1,473
2025-10	h3	105,2925	107,38	105,2272	2,1528
2025-11	h3	105,7001	108,1319	105,6347	2,4972
2025-07	h6	105,0399	107,38	105,4054	1,9746
2025-08	h6	104,6479	108,1319	105,1525	2,9794

Fuente: Autores

El error absoluto promedio aumento conforme se amplió el horizonte de predicción. En h1 el MAE fue 0.4544, mientras que en h6 fue 2.4770. Por otro lado, El error absoluto promedio aumento conforme se amplió el horizonte de predicción. En h1 el MAE fue 0.4544, mientras que en h6 fue 2.4770 como se muestra en la Tabla 19.

Tabla 19
Métricas por cada horizonte

Horizonte	Observaciones	Media	Desviación estándar	Mínimo	Mediana	Máximo
h1	7	0,4544	0,1923	0,1762	0,53	0,6648
h3	5	1,5658	0,7951	0,4711	1,473	2,4972

h6	2	2,477	0,7105	1,9746	2,477	2,9794
----	---	-------	--------	--------	-------	--------

Fuente: Autores

4.2.6.7 Comparación con Naive Persistence

A continuación, en la Tabla 20 se compara LightGBM con el modelo Naive Persistence. Esto permite evaluar si aporta una mejora frente a la línea base.

Tabla 20
Comparación LightGBM con Naive Persistence

Horizonte	MAE Naive	MAE LightGBM	Variación frente a Naive
h1	0,5576	0,4544	18,51% menor
h3	1,4229	1,5658	10,04% mayor
h6	2,912	2,477	14,94% menor

Fuente: Autores

De la tabla anterior, se observa que el modelo LightGBM mejoró el MAE en h1 y h6. En h3, el MAE fue mayor que el de Naive Persistence. Sugiriendo así, que la incorporación de variables y la transformación del objetivo no mejoran de forma uniforme todos los horizontes.

Por otro lado, La importancia de variables se calculó con la ganancia generada por el modelo de LightGBM para la mejor configuración de cada horizonte. Los resultados muestran que el ITCER y sus rezagos son muy relevantes, y también las variables macroeconómicas aportan información en algunos horizontes, el h1 y h6.

Variables principales segun ganancia:

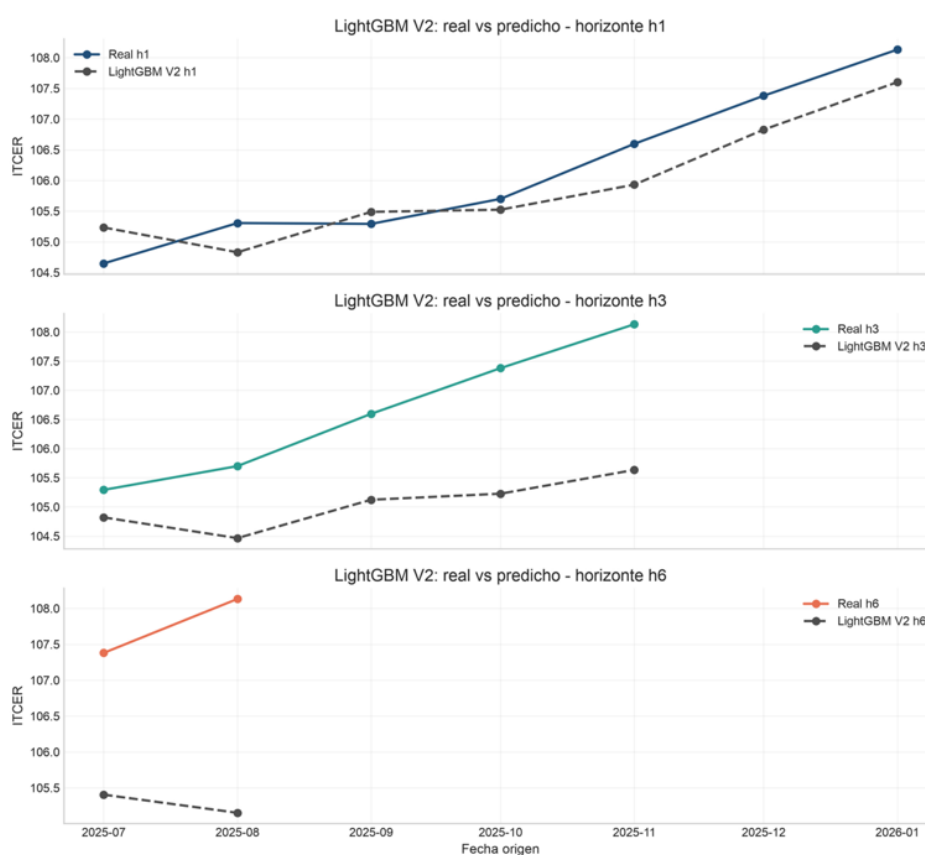
- **h1:** mes, ITCER_lag1, ITCER_lag3, ITCER, USD_NEER_lag1, EMBI
- **h3:** ITCER, tendencia, ITCER_lag3, USD_NEER_lag1, mes_cos, EMBI_lag1
- **h6:** ITCER, mes, WTI_lag1, WTI, USD_NEER, EMBI_lag1

4.2.6.8 Resultados

La Figura17 muestra la comparación consolidada del valor real vs el valor predicho para h1 h3 h6. Donde se observa que el modelo LightGBM sigue de forma cercana los valores de h1, pero la diferencia aumenta en h3 y h6. En los horizontes más largos, el modelo tiende a quedar por debajo del valor observado, especialmente al final del periodo de prueba.

Figura 17

Comparación entre los valores reales y predichos del modelo LightGBM para los horizontes h1, h3 y h6



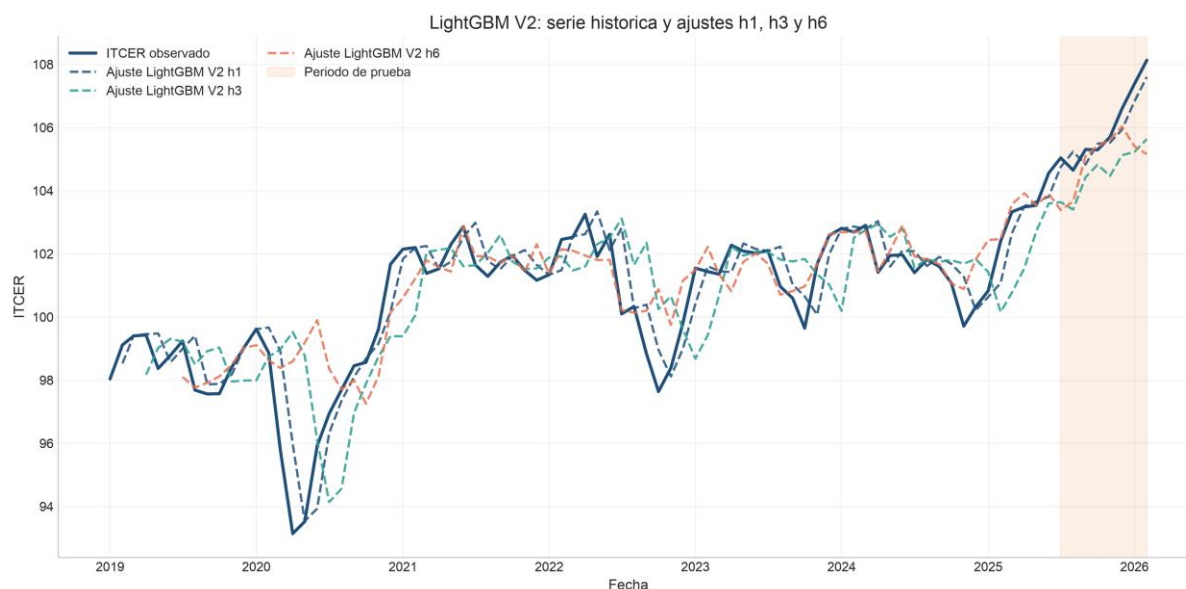
Fuente: Autores

Por otro lado, tenemos la Figura18 para la serie histórica, donde se observa que los ajustes del modelo capturan parte del movimiento general del ITCER. Sin embargo, en el

tramo de prueba el crecimiento de la serie es más marcado que el ajuste estimado para algunos horizontes.

Figura 18

Serie histórica del ITCER y ajustes del modelo LightGBM para los horizontes h1, h3 y h6



Fuente: Autores

4.2.7 Estimación modelo LSTM multi-horizonte

El modelo Long Short-Term Memory (LSTM) es una arquitectura de redes neuronales recurrentes diseñada para modelar dependencias temporales en datos secuenciales. Las LSTM incorporan una estructura de memoria que les permite conservar información relevante de observaciones anteriores y descartar aquella que no aporta al proceso de predicción.

En esta investigación se implementó un modelo LSTM multi-horizonte, cuyo objetivo es predecir simultáneamente varios valores futuros del Índice de Tipo de Cambio Real Efectivo (ITCER) a partir de una secuencia de observaciones históricas y variables macroeconómicas relacionadas.

Con el objetivo de evaluar el potencial de las redes neuronales para el pronóstico del ITCER, se implementaron diferentes arquitecturas basadas en Long Short-Term Memory (LSTM). Estas arquitecturas fueron diseñadas para realizar pronósticos multi-horizonte utilizando la información histórica del ITCER y un conjunto de variables macroeconómicas seleccionadas.

Previo al entrenamiento, todas las variables fueron normalizadas mediante el algoritmo MinMaxScaler, con el fin de llevarlas al intervalo $[0,1]$ y facilitar el proceso de aprendizaje de la red neuronal. Posteriormente, los datos fueron organizados en secuencias temporales mediante ventanas deslizantes (*sliding windows*), permitiendo que el modelo utilizara observaciones históricas para estimar varios períodos futuros de forma simultánea.

Durante el entrenamiento se empleó el optimizador Adam y la función de pérdida Mean Squared Error (MSE). Asimismo, se incorporó la técnica Early Stopping, la cual permitió detener automáticamente el entrenamiento cuando el error de validación dejó de mejorar, reduciendo el riesgo de sobreajuste.

Con el objetivo de identificar la arquitectura con mejor capacidad predictiva, se evaluaron tres configuraciones diferentes del modelo LSTM, se resumen en la Tabla 21:

Tabla 21
Arquitecturas evaluadas del modelo LSTM multi-horizonte

Modelo	N_STEPS_IN	N_STEPS_OUT	Arquitectura
LSTM Multi-Horizonte Original	12	6	LSTM(100, relu) -> Dropout(0.2) -> Dense(6)
LSTM 3 Meses	5	3	LSTM(64, relu) -> Dropout(0.2) -> Dense(32, relu) -> Dense(3)
LSTM 6 Meses Ajustado	5	6	LSTM(64, relu, return_seq=True) -> Dropout(0.2) -> LSTM(64, relu) -

> Dropout(0.2) -> Dense(32, relu) -
> Dense(6)

Fuente: Autores

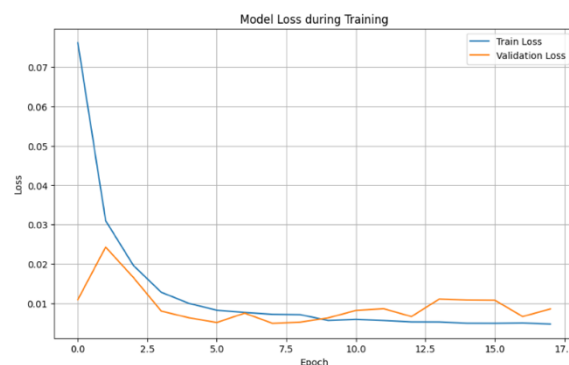
Las arquitecturas difieren principalmente en el número de observaciones utilizadas como entrada, el horizonte de predicción y la complejidad de la red neuronal. Inicialmente se implementó una arquitectura con una única capa LSTM y una ventana temporal de doce meses.

Posteriormente, se evaluaron arquitecturas más compactas utilizando ventanas de tres y seis meses y una versión de mayor profundidad compuesta por dos capas LSTM, con el propósito de determinar si una arquitectura más compleja mejoraba la capacidad predictiva del modelo, como se observa en la Figura 19.

En la Figura 19, 20 y 21, el comportamiento de las funciones de pérdida evidencia que los modelos convergen durante el proceso de entrenamiento sin presentar inestabilidades significativas. La implementación de Early Stopping permitió conservar la configuración con menor error de validación, evitando el sobreajuste y favoreciendo una mejor capacidad de generalización sobre datos no observados.

Figura 19

Función de perdida de modelo LSTM 1



Fuente: Autores

Figura 20
Función de pérdida de modelo LSTM 2

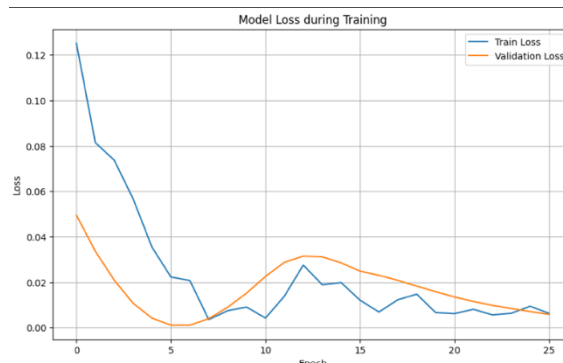
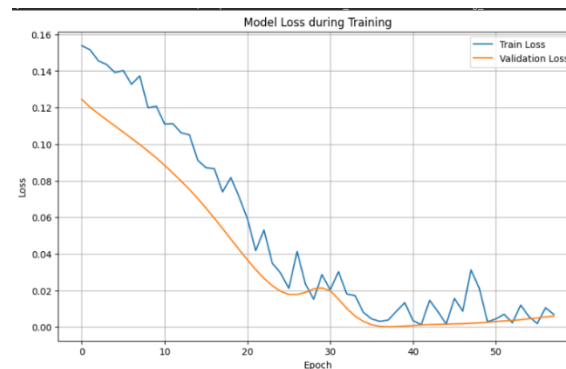


Figura 21
Función de pérdida de modelo LSTM 3



Fuente: Autores

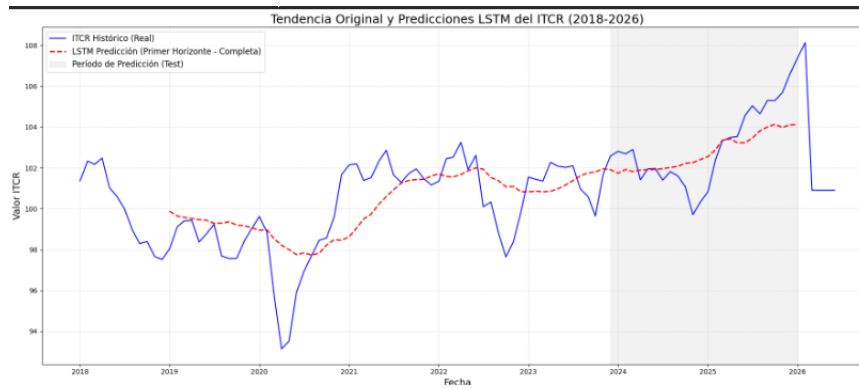
Los resultados muestran diferencias importantes entre las arquitecturas evaluadas. El modelo LSTM Multi-Horizonte Original, construido con una ventana de doce meses y un horizonte de predicción de seis períodos, presentó un seguimiento más cercano del comportamiento observado del ITCER. En contraste, las arquitecturas de tres y seis meses ajustados mostraron una mayor dispersión en las predicciones, especialmente cuando la serie presentó cambios más pronunciados.

Aunque las arquitecturas más profundas incrementan la capacidad de aprendizaje de la red, los resultados obtenidos indican que dicho aumento de complejidad no necesariamente se traduce en una mejora del desempeño predictivo. Considerando que el conjunto de datos utilizado contiene un número reducido de observaciones, una arquitectura excesivamente

compleja puede favorecer el sobreajuste y disminuir la capacidad de generalización del modelo.

Figura 22

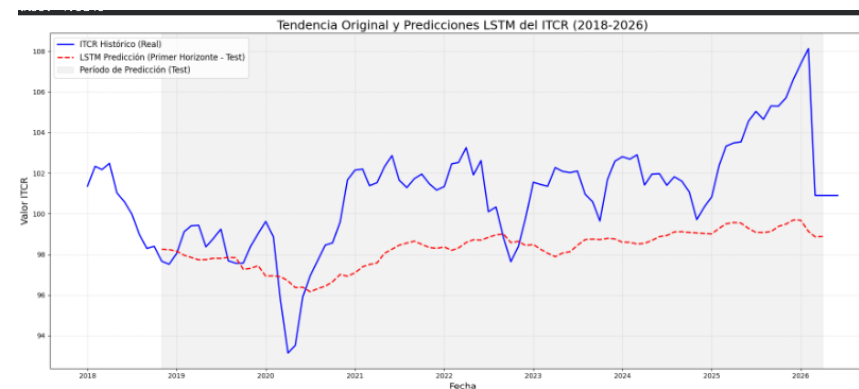
Pronósticos obtenidos mediante los modelos LSTM multi-horizonte: Modelo 1



Fuente: Autores

Figura 23

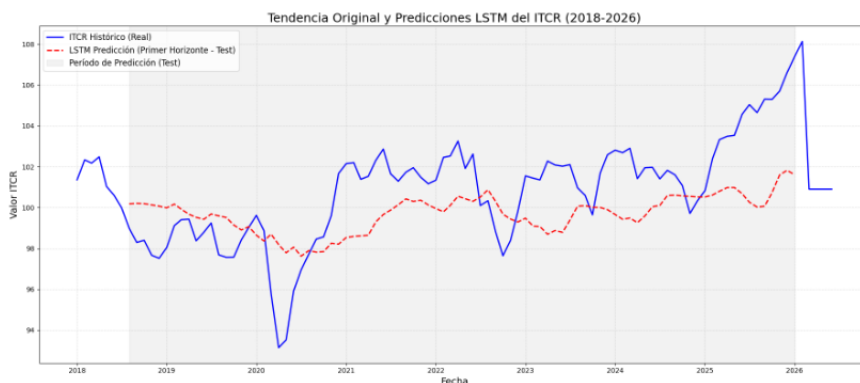
Pronósticos obtenidos mediante los modelos LSTM multi-horizonte: Modelo 2



Fuente: Autores

Figura 24

Pronósticos obtenidos mediante los modelos LSTM multi-horizonte: Modelo 2



En la Figura 22, 23 y 24 los resultados obtenidos muestran que el modelo LSTM multi-horizonte logra capturar adecuadamente la dinámica temporal del ITCER, aprovechando tanto la información histórica de la serie como la influencia de las variables macroeconómicas incorporadas. El desempeño cuantitativo del modelo se analiza en la siguiente sección mediante las métricas de evaluación correspondientes.

4.2.7.1 Evaluación del desempeño del modelo

Con el propósito de comparar objetivamente las arquitecturas implementadas, se calcularon diferentes métricas de evaluación sobre el conjunto de prueba. En particular, se emplearon el Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE), el Error Absoluto Porcentual Medio (MAPE) y el coeficiente de determinación (R^2). Dichos valores se adjuntan en la Tabla 22.

Tabla 22

Métricas de desempeño de los modelos LSTM multi-horizonte

Modelo	RMSE	MAE	MAPE	R-squared
LSTM Multi-Horizonte Original	2.06	1.68	1.62%	0.12
LSTM 3 Meses	3.65	3.05	N/A	N/A
LSTM 6 Meses Ajustado	3.65	3.05	N/A	N/A

Fuente: Autores

La Tabla 22 muestra que la arquitectura LSTM Multi-Horizonte Original obtuvo el mejor desempeño entre los modelos evaluados, alcanzando un RMSE de 2.06, un MAE de 1.68, un MAPE de 1.62 % y un coeficiente de determinación de 0.12.

Por su parte, las arquitecturas diseñadas para horizontes de tres y seis meses registraron errores superiores, con valores de RMSE de 3.65 y MAE de 3.05, lo que evidencia una menor capacidad para reproducir el comportamiento del ITCER durante el período de evaluación.

En términos generales, los resultados indican que una arquitectura relativamente simple logró un mejor equilibrio entre capacidad de aprendizaje y generalización que las configuraciones de mayor complejidad. No obstante, el valor obtenido del coeficiente de determinación muestra que el modelo aún presenta limitaciones para explicar completamente la variabilidad del ITCER, lo que justifica la evaluación de arquitecturas más avanzadas basadas en mecanismos de atención, como el Temporal Fusion Transformer, desarrollado en la siguiente sección.

4.2.8 Estimación Modelo TFT (Temporal Fusion Transformer)

El modelo Temporal Fusion Transformer (TFT) se incorporó como una alternativa de aprendizaje profundo para el pronóstico mensual del ITCER. Su objetivo fue evaluar si una arquitectura con selección de variables, memoria temporal recurrente y atención puede aprovechar mejor la información histórica del índice y de algunas variables macroeconómicas.

A diferencia de otros modelos del proyecto, este modelo se entrenó para pronosticar directamente el nivel futuro del ITCER en varios horizontes. La salida del modelo corresponde a una trayectoria de hasta seis meses hacia adelante:

$$\widehat{y}_{t+1}, \widehat{y}_{t+2}, \dots, \widehat{y}_{t+6}$$

Por otro lado, El TFT es una arquitectura diseñada para problemas de pronóstico con mucha disponibilidad de datos, varias series temporales, variables estáticas, covariables históricas, covariables futuras conocidas y relaciones no lineales muy amplias. Es decir, en estos casos, el modelo pueda aprovechar de mejor manera su estructura completa: redes de selección de variables, capas recurrentes, bloques de compuertas, redes residuales y mecanismos de atención.

En el presente trabajo, sobre pronóstico de ITCER del Ecuador, la información disponible tiene una limitación importante: la serie es mensual y cubre el periodo de enero de 2018 a febrero de 2026. Aunque el intervalo temporal es relevante para cualquier análisis económico, el número de observaciones es pequeño para una red neuronal profunda. Al trabajar con una ventana histórica de 12 meses y un horizonte máximo de 6 meses, el número real de uso de ventanas de entrenamiento llega a disminuir mucho.

Por esta razón, se construyó una versión disminuida del TFT. Se buscó no replicar toda la complejidad de la arquitectura original, sino conservar sus componentes principales y compatible con el tamaño de la muestra de este proyecto. En particular, se mantuvieron tres elementos:

- Selección de variables históricas;
- Representación temporal mediante LSTM;
- Atención temporal sobre la ventana histórica.

También, se redujo el tamaño de las capas internas, el número de cabezas de atención y el número de capas recurrentes. Esta configuración tuvo como objetivo disminuir la cantidad de parámetros entrenables y reducir el riesgo de sobreajuste. En una muestra mensual pequeña como la nuestra, una arquitectura muchos parámetros puede ajustar ruido en el entrenamiento y perder capacidad de generalización fuera de muestra.

El ajuste de hiperparámetros se realizó priorizando la estabilidad del entrenamiento y la reducción del sobreajuste. Se utilizó una tasa de aprendizaje de 0.001 con el optimizador AdamW, que incorpora decaimiento de pesos mediante $\text{weight_decay} = 0.0001$. Además, se aplicó $\text{dropout} = 0.10$ para reducir la dependencia del modelo respecto a combinaciones específicas de neuronas, y $\text{clipnorm} = 0.1$ para controlar posibles variaciones bruscas en los gradientes durante el entrenamiento. La función de pérdida seleccionada fue Huber, ya que permite combinar la sensibilidad del error cuadrático con una menor exposición a valores extremos.

El entrenamiento se limitó a un máximo de 300 épocas, utilizando EarlyStopping como mecanismo de detención automática. La razón fue evitar entrenamientos muy largos y reducir el riesgo de que el modelo aprendiera patrones específicos del conjunto de entrenamiento. Finalmente, el modelo se evaluó con varias semillas de inicialización, específicamente 7, 8, 9, 10 y 11, para revisar la sensibilidad de los resultados frente a la aleatoriedad propia del entrenamiento de redes neuronales.

4.2.8.1 Datos utilizados y partición temporal

La partición temporal se mantuvo alineada con los demás modelos del proyecto. El entrenamiento final utilizó la información disponible hasta junio de 2025, mientras que la prueba se realiza desde julio de 2025 hasta febrero de 2026.

Dado que el modelo requiere una ventana histórica de 12 meses, la primera fecha de origen real aparece cuando ya existe suficiente información previa para formar dicha ventana. Por ello, aunque la base comienza en enero de 2018, las ventanas de modelado inician efectivamente en diciembre de 2018.

- Entrenamiento interno: 2018-12 a 2024-06
- Validación: 2025-01 a 2025-06

- Entrenamiento final: 2018-12 a 2025-06
- Prueba: 2025-07 a 2026-02

La validación se utilizó para controlar el entrenamiento y seleccionar el número de épocas. Luego, el modelo final se reentreno con todas las ventanas disponibles hasta junio de 2025. De esta forma, la evaluación de prueba mantiene el mismo criterio temporal de los otros modelos.

La cantidad de observaciones reales disponibles para evaluar disminuye a medida que aumenta el horizonte. En h1 se cuenta con siete comparaciones, en h3 con cinco y en h6 con dos. Esta diferencia debe considerarse al interpretar los resultados, especialmente en h6.

4.2.8.2 Variables utilizadas^[Obj]

Para determinar la utilidad de las variables exógenas, se compararon dos instancias principales: una univariada y otra multivariada. La instancia univariada utiliza solo la historia del ITCER. La instancia multivariada incorpora las variables macroeconómicas observadas en la ventana histórica.

- TFT univariado: ITCER, mes_sin, mes_cos
- TFT multivariado: ITCER, IPC_ECU, IPC_USA, USD_NEER, WTI, EMBI, mes_sin, mes_cos.

Las variables `mes\_sin` y `mes\_cos` representan la estacionalidad mensual mediante una tendencia cíclica, permitiendo que el modelo trate la variable calendario como un ciclo continuo. Por ejemplo, diciembre y enero quedan cercanos dentro de la representación, lo cual no ocurre si el mes se ingresa como un numero entero de 1 a 12. Por otro lado, no se incluyeron rezagos explícitos como variables adicionales, porque la ventana histórica de 12 meses ya entrega al modelo la información rezagada, es decir, agregar rezagos calculados sobre una ventana temporal aumentaría la dimensionalidad sin aportar información nueva

para el modelado. Tampoco se utilizaron valores futuros de variables macroeconómicas, ya que esos valores no están disponibles al momento de generar el pronóstico. Es decir, incluirlos equivaldría a introducir información futura en el modelo. Por esa razón, las variables futuras conocidas se limitaron al calendario.

4.2.8.3 Adecuación de enfoques univariados y multivariados^[OBJ]

El uso de un modelo univariado y multivariado tiene el propósito de comparar el desempeño del modelo. Comprobar si la propia historia de la variable objetivo contiene la mayor parte de la información predictiva. El enfoque de modelar solo la serie objetivo es para tener una referencia, y evitar incorporar variables adicionales que pueden introducir ruido, inestabilidad o problemas de disponibilidad futura. Sin embargo, también puede ser limitado si los movimientos del ITCER están asociados con información externa, como precios relativos, condiciones financieras o movimientos del tipo de cambio nominal.

Por otro lado, la instancia multivariada busca incorporar ese contexto adicional. En este proyecto se incluyeron variables relacionadas con precios internos y externos, condiciones externas y riesgo país.

Sin embargo, usar más variables no garantiza mejor desempeño. En nuestro modelo reducido, cada variable adicional aumenta la complejidad del problema. Si la relación entre las variables explicativas y el ITCER no es significativa, el modelo puede aprender asociaciones específicas del periodo de entrenamiento que no se repiten en la muestra de prueba. Por ello, se eligió un conjunto acotado de variables macroeconómicas, en lugar de ampliar la base con muchas series adicionales.

4.2.8.4 Configuración del modelo

El modelo se implementó con TensorFlow y Keras. Se utilizó una versión liviana del TFT, con una dimensión oculta reducida y una sola cabeza de atención. El objetivo fue

conservar la estructura interpretativa del modelo sin aumentar mucho la cantidad de parámetros. Los hiperparámetros utilizados se muestran en la Tabla 23.

Tabla 23
Configuración del modelo TFT

Elemento	Configuración
Ventana histórica	12 meses
Horizonte máximo	6 meses
Salidas evaluadas	h1, h3 y h6
Función de pérdida	Huber
Optimizador	AdamW
Tasa de aprendizaje	0,001
Dimensión oculta	8
Dimensión continua interna	4
Capas LSTM	1
Cabezas de atención	1
Dropout	0,1
Batch size	8
Épocas máximas	300
Control de entrenamiento	EarlyStopping
Regularización	weight decay = 0,0001
Control de gradiente	clipnorm = 0,1
Semillas evaluadas	7, 8, 9, 10 y 11

Fuente: Autores

La función de pérdida Huber se eligió porque combina las propiedades del error cuadrático y del error absoluto. Penaliza de forma suave los errores pequeños y reduce la sensibilidad frente a errores grandes. Esta característica llega a ser útil cuando el periodo de prueba tiene pocos datos. El optimizador AdamW corresponde a una variante de Adam que

incorpora decaimiento de pesos. Su uso permite controlar el crecimiento de los parámetros durante el entrenamiento.

4.2.8.5 Parámetros entrenables

En las redes neuronales, los parámetros que se entrenan son los pesos y los sesgos que el modelo ajusta durante el entrenamiento. Estos parámetros determinan como se combinan las variables de entrada, como se transforma la información en las capas internas y como se producen las predicciones finales.

La Tabla 24 muestra el tamaño aproximado de las dos instancias:

Tabla 24
Parámetros entrenables de TFT

Modelo	Parámetros entrenables	Ventanas del modelo final
TFT univariado	1961	79
TFT multivariado	2041	79

Fuente: Autores

Aunque la arquitectura fue reducida, la relación entre parámetros y ventanas es bastante alta. Esto no significa que el modelo este mal configurado, pero si indica que el resultado del aprendizaje debe usarse con cuidado. En modelos de aprendizaje profundo, una cantidad limitada de observaciones puede dificultar la estimación estable de los parámetros.

4.2.8.6 Funcionamiento general de la arquitectura

El TFT utilizado en este proyecto sigue una lógica de procesamiento en tres etapas. Primero, el modelo recibe la ventana histórica de 12 meses. Para el caso multivariado, esta ventana contiene el ITCER y las variables macroeconómicas seleccionadas. En el caso univariado, contiene solo el ITCER.

Segundo, el modelo aplica redes de selección de variables. Donde estas redes asignan pesos internos a las variables de entrada con el objetivo de permitir que el modelo determine que variables son más significante para representar la información histórica. En la práctica, estos pesos se calculan dentro de la red y cambian según el patrón de entrada.

Tercero, la información seleccionada pasa por capas LSTM y por un mecanismo de atención temporal. La LSTM resume la evolución de la ventana histórica, mientras que la atención permite que el modelo le asigne mayor o menor importancia a determinados tiempos previos al generar los pronósticos. Finalmente, una capa de salida produce las predicciones para los seis meses futuros.

4.2.8.7 Interpretabilidad del TFT

Una de las razones para evaluar la implementación de TFT es que, a diferencia de otras redes neuronales, incluye mecanismos que ayudan a indentificar como el modelo organiza la información temporal como de las variables input . Para ello se utilizaron cuatro criterios. Primero, se analizo la importancia de variables historicas a partir de los pesos generados por la red de seleccion de variables del encoder. Segundo, se reviso la importancia de las variables futuras conocidas del decoder, que en este caso corresponden a las variables de calendario `mes\_sin` y `mes\_cos`. Tercero, se evaluo la atencion temporal para determinar si el modelo concentraba sus predicciones en rezagos especificos dentro de la ventana historica. Cuarto, se realizo un analisis de ablacion (eliminación, supresión), retirando una variable exogena a la vez, para observar si el conjunto multivariado aportaba informacion frente a versiones con menos varibales del TFT.

4.2.8.8 Importancia de variables históricas

La importancia de variables proviene de la red de selección del encoder. Este encoder recibe la ventana histórica y calcula pesos para las variables disponibles. En la instancia

multivariada, estos pesos permiten revisar que variables tienen mayor importancia y participación dentro del modelo. En el modelo multivariado seleccionado, la importancia promedio que generó el encoder se muestra en la Tabla 25.

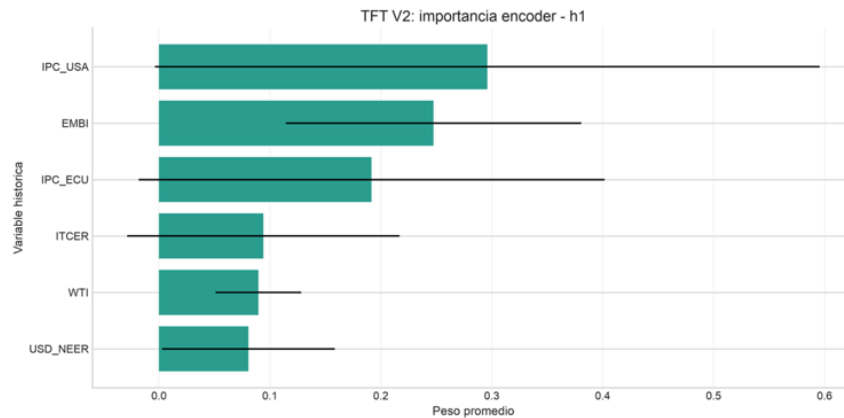
Tabla 25
Importancia de variables históricas

Variable	Importancia promedio	Desviación estándar
IPC_USA	0,296	0,2996
EMBI	0,2476	0,1333
IPC_ECU	0,1917	0,21
ITCER	0,0942	0,1228
WTI	0,0897	0,0386
USD_NEER	0,0807	0,0779

Fuente: Autores

En el TFT Light multivariado, las variables con mayor peso promedio en el encoder fueron IPC_USA, EMBI e IPC_ECU. Esto muestra que el modelo dio más importancia a variables relacionadas con precios externos, riesgo país y precios internos. Sin embargo, hay que considerar que la importancia del encoder no mide causalidad entre las variables. Solo describe como el modelo pondera las variables dentro de su arquitectura respecto a su generación de predicciones.

Figura 25
Encoder de TFT



Fuente: Autores

La Figura 25 sobre la importancia del encoder muestra que el modelo asignó mayor peso relativo a variables como IPC_USA, EMBI e IPC_ECU. Estos pesos describen el uso interno de las variables, no efectos causales.

4.2.8.9 Importancia de variables futuras conocidas

La importancia que da el decoder se calculó sobre las variables futuras conocidas. En este caso, las variables futuras que se incluyeron fueron las de `mes\_sin` y `mes\_cos`, que representan el calendario, los valores se detallan en la Tabla 26.

Tabla 26
Comparativa horizontes 1, 3 y 6 meses

Horizonte	mes sin	mes cos
h1	0,4555	0,5445
h3	0,4895	0,5105
h6	0,5407	0,4593

Fuente: Autores

Los pesos muestran que el modelo utilizó ambas componentes del calendario, con diferencias pequeñas entre horizontes. Esto sugiere que el calendario entró como información de apoyo, pero no permite concluir que exista un patrón estacional fuerte.

4.2.8.10 Atención temporal

La atención temporal permite indicar que meses de la ventana histórica reciben mayor peso al generar los pronósticos. Es decir, si el modelo concentrara la atención en unos pocos rezagos, se podría inferir que ciertos meses pasados fueron más relevantes que otros en la predicción. En la siguiente Tabla 27 podemos ver esos pesos:

Tabla 27
Atención temporal

Modelo	Horizonte	Posición histórica	Peso promedio	Peso std	Peso mín.	Peso máx.
TFT multivariado	h1	t	0.0832	0.0053	0.0750	0.0894
TFT multivariado	h1	t-1	0.0829	0.0050	0.0751	0.0890
TFT multivariado	h1	t-10	0.0856	0.0085	0.0755	0.0980
TFT multivariado	h1	t-11	0.0890	0.0150	0.0736	0.1097
TFT multivariado	h1	t-2	0.0826	0.0047	0.0754	0.0885
TFT multivariado	h1	t-3	0.0823	0.0044	0.0758	0.0878
TFT multivariado	h1	t-4	0.0821	0.0041	0.0766	0.0869
TFT multivariado	h1	t-5	0.0820	0.0035	0.0777	0.0856
TFT multivariado	h1	t-6	0.0819	0.0029	0.0783	0.0846
TFT multivariado	h1	t-7	0.0821	0.0022	0.0788	0.0847
TFT multivariado	h1	t-8	0.0826	0.0023	0.0799	0.0852
TFT multivariado	h1	t-9	0.0837	0.0045	0.0777	0.0903
TFT multivariado	h3	t	0.0829	0.0048	0.0749	0.0872
TFT multivariado	h3	t-1	0.0826	0.0046	0.0751	0.0870
TFT multivariado	h3	t-10	0.0863	0.0081	0.0784	0.0981
TFT multivariado	h3	t-11	0.0901	0.0149	0.0772	0.1100

Modelo	Horizonte	Posición histórica	Peso promedio	Peso std	Peso mín.	Peso máx.
TFT multivariado	h3	t-2	0.0822	0.0045	0.0753	0.0867
TFT multivariado	h3	t-3	0.0819	0.0043	0.0758	0.0863
TFT multivariado	h3	t-4	0.0818	0.0040	0.0765	0.0856
TFT multivariado	h3	t-5	0.0817	0.0035	0.0777	0.0847
TFT multivariado	h3	t-6	0.0817	0.0030	0.0778	0.0844
TFT multivariado	h3	t-7	0.0820	0.0023	0.0785	0.0848

Fuente: Autores

Como se observa en los resultados obtenidos, los pesos de atención fueron similares a lo largo de la ventana de 12 meses. Pero, se observa una ponderación ligeramente mayor en los rezagos más lejanos, especialmente alrededor de t-11 y t-10, pero la diferencia frente al resto de meses fue pequeña.

Esta distribución casi uniforme sugiere que el modelo no identificó un punto o región temporal dominante dentro de la ventana histórica. En nuestro caso de poca historia temporal, esto puede reflejar que el mecanismo de atención no tuvo información suficiente para aprender patrones temporales.

4.2.8.11 Experimento de selección de variables exógenas (ablación)

Además de comparar el modelo univariado y el multivariado, se realizó un análisis de selección de variables exógenas. Este ejercicio consiste en retirar una variable exógena a la vez y observar cómo cambia el ajuste del modelo. Su objetivo es revisar si el modelo depende de una variable específica o si el conjunto completo aporta mejor información.

Tabla 28
Ranking del MAE promedio global

Modelo	MAE promedio global	Desv. MAE global
--------	---------------------	------------------

TFT multivariado	4,0479	0,8109
TFT sin IPC_ECU	4,3256	1,458
TFT sin EMBI	4,4523	0,7569
TFT sin IPC_USA	4,522	1,3856
TFT sin WTI	4,6498	0,9032
TFT sin USD_NEER	4,9391	0,5491
TFT univariado	5,5407	1,0711

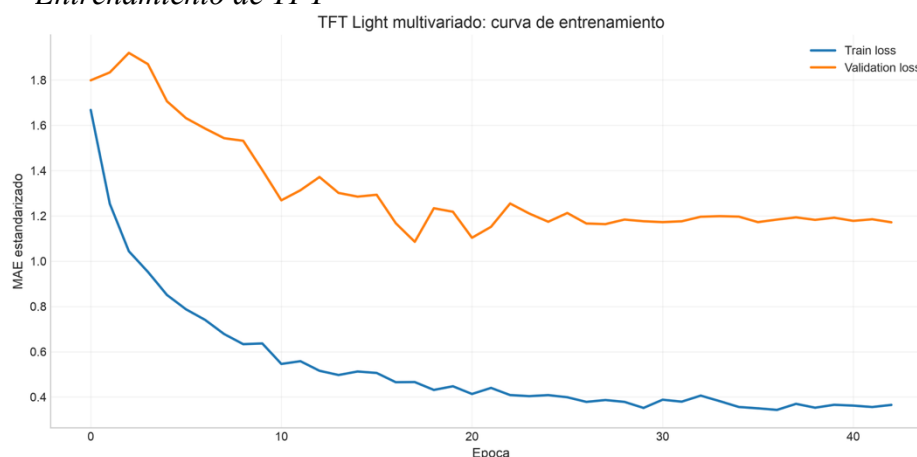
Fuente: Autores

Como se puede observar en la Tabla 28, el modelo multivariado completo obtuvo el menor MAE promedio global. Esto sugiere que para el modelo TFT el conjunto completo de variables históricas aportó más información que las versiones reducidas, con menos variables exógenas.

A continuación, se muestra la curva de entrenamiento:

Figura 26

Entrenamiento de TFT



La Figura 26 muestra la evolución de la pérdida durante el entrenamiento del modelo seleccionado, donde, la validación se utilizó para controlar el entrenamiento.

4.2.8.12 Predicciones del modelo seleccionado

Para las gráficas y el pronóstico final se seleccionó el modelo TFT multivariado. La Tabla 29 muestra las predicciones fuera de muestra disponibles.

Tabla 29
Comparativa entre predicciones TFT

Fecha de origen	ITCER observado	Real h1	Pred. h1	Real h3	Pred. h3	Real h6	Pred. h6
2025-07	105,0399	104,6479	102,2227	105,2925	102,2565	107,38	102,9818
2025-08	104,6479	105,306	102,167	105,7001	102,4949	108,1319	103,1779
2025-09	105,306	105,2925	102,0328	106,5963	102,5972	-	103,2164
2025-10	105,2925	105,7001	102,1369	107,38	102,6982	-	103,2591
2025-11	105,7001	106,5963	102,2788	108,1319	102,8299	-	103,2248
2025-12	106,5963	107,38	102,4534	-	102,9958	-	103,1688
2026-01	107,38	108,1319	102,6356	-	103,1091	-	103,0556
2026-02	108,1319	-	102,5474	-	102,8958	-	103,0392

Fuente: Autores

Las predicciones muestran una subestimación continua durante el periodo de prueba. El modelo mantuvo los pronósticos en el rango de 102-103, mientras que el ITCER observado aumento durante los meses evaluados.

Aunque el TFT no tuvo un buen desempeño, este resultado sigue siendo valioso para el trabajo. La evidencia muestra que, para el ITCER mensual de Ecuador, un modelo con atención y selección de variables necesita más datos, más series relacionadas o variables futuras conocidas para dar mejores pronósticos. En este caso, el modelo subestimo los valores, predijo valores entre 102 y 103 puntos, mientras que el ITCER real subió a cerca de 108. Por eso, la comparación deja ver que usar un modelo más complejo tiene que justificarse con resultados y no siempre va a funcionar mejor.

4.2.8.13 Comparación con otros modelos del proyecto

La Tabla 30 resume el desempeño del TFT frente a modelos de referencia y frente a otros modelos desarrollados en el proyecto.

Tabla 30
Comparativa entre modelos

Modelo	MAE h1	MAE h3	MAE h6
Naive Persistence	0,5576	1,4229	2,912
Seasonal Naive	5,0422	5,7523	6,1417
LightGBM directo multi-horizonte	0,4544	1,5658	2,477
TiDE multivariado	0,4614	0,9718	1,725
TiDE univariado	0,4902	1,484	1,506
TFT multivariado	3,4086	3,9193	4,8159
TFT univariado	4,9696	5,4464	6,2061

Fuente: Autores

Como se observa, el TFT multivariado no logro superar a Naive Persistence. En h1, su MAE fue aproximadamente seis veces mayor que el de la línea base de persistencia. En h3 y h6 también registro errores superiores. Esta diferencia es relevante porque Naive Persistence es un modelo simple que proyecta el ultimo valor observado.

4.2.8.14 Pronostico consolidado h1, h3 y h6

A continuación, la Figura 27 consolida los pronósticos de los tres horizontes evaluados. Donde, podemos observar que los pronósticos del modelo se mantienen relativamente planos frente al aumento del ITCER observado.



4.2.9 Estimación Modelo TiDE

El modelo TiDE fue desarrollado como una alternativa en los modelos de aprendizaje profundo para el pronóstico del ITCER. Su objetivo es evaluar si una arquitectura temporal densa puede mejorar el desempeño frente a modelos base y frente a los modelos supervisados más simples.

En este modelo no se predice el nivel futuro directamente del ITCER. Primero se estima el cambio esperado respecto al valor observado en la fecha de origen y luego se reconstruye el nivel pronosticado. La transformación del objetivo fue:

$$target_{t,h} = y_{t+h} - y_t$$

El pronóstico final se obtuvo como:

$$\widehat{y_{t+h}} = y_t + \widehat{target_{t,h}}$$

Donde:

- y_t : ITCER observado en la fecha de origen
- y_{t+h} : ITCER observado en el horizonte futuro
- $target_{t,h}$: Cambio futuro del ITCER respecto al valor actual
- $\widehat{y_{t+h}}$: Pronostico final del nivel del ITCER
- h Horizonte de predicción: 1, 3 o 6 meses

El modelo se estimó con la base mensual del ITCER. La partición temporal se mantuvo con entrenamiento hasta junio de 2025 y prueba desde julio de 2025 hasta febrero de 2026.

Por otro lado, la arquitectura trabaja con ventanas históricas de 12 meses y pronósticos hasta 6 meses hacia adelante. Por esa razón, la fecha de origen efectiva comienza cuando existe suficiente historia para formar la primera ventana. Quedando finalmente como:

- 2018-01 a 2024-06: se usa para entrenar pesos.
- 2025-01 a 2025-06: se usa como validación interna. Pero, para el modelo final se usan hasta el 2025-06.
- 2025-07 a 2026-02: se usa como test.

Para el modelo TiDE el modelamiento se organizó en dos etapas. Primero, se utilizó una validación interna dentro del periodo de entrenamiento para definir el numero adecuado de épocas y controlar el ajuste del modelo. Para esta etapa, el entrenamiento uso las ventanas disponibles hasta junio de 2024, mientras que el periodo enero-junio de 2025 se empleó como validación. Esta validación no formó parte del conjunto de prueba.

El segundo paso, después definida la configuración de entrenamiento, se ajustó el modelo final usando todos los datos hasta junio de 2025. Así, el modelo evaluado en prueba fue entrenado con el mismo periodo utilizado para los demás modelos del proyecto, y manteniendo el conjunto de prueba entre julio de 2025 y febrero de 2026.

De manera general, el modelo se configuró como una arquitectura temporal densa de baja complejidad. Se compararon una instancia univariada, basada solo en el ITCER, y una instancia multivariada, que incluyo ITCER, IPC de Ecuador, IPC de Estados Unidos, USD_NEER, WTI y EMBI. Para el tramo futuro se incorporaron únicamente variables de calendario codificadas como mes_sin y mes_cos.

4.2.9.1 Variables utilizadas

Se compararon dos instancias diferentes. La primera solo usa la historia del ITCER. La segunda incorpora variables macroeconómicas observadas en la ventana histórica. En ambos casos, las variables futuras conocidas se limitaron a componentes de calendario. Las dos instancias son:

- TiDE univariado: ITCER, mes_sin, mes_cos.

- TiDE multivariado: ITCER, IPC_ECU, IPC_USA, USD_NEER, WTI, EMBI, mes_sin, mes_cos.

Las variables `mes\_sin` y `mes\_cos` representan la estacionalidad mensual mediante una codificación cíclica. Esta forma evita tratar enero y diciembre como meses lejanos cuando, en términos de calendario, son meses consecutivos.

El modelo se implementó con TensorFlow y Keras. Se utilizó una versión con estructura y parámetros conservadores de TiDE, basada en bloques densos residuales, sin capas recurrentes, sin mecanismos de atención y sin variables exógenas futuras distintas al calendario.

Tabla 31
Parámetros de entrenamiento TiDE

Elemento	Configuración
Ventana histórica	12 meses
Horizonte máximo	6 meses
Salidas evaluadas	h1, h3 y h6
Función de pérdida	Huber
Optimizador	AdamW
Dimensión oculta	16
Dimensión del decodificador	8
Dropout	0,1
Épocas máximas	500
Control de entrenamiento	EarlyStopping

Fuente: Autores

Como se observa en la Tabla 3 la configuración se mantuvo reducida para evitar una cantidad muy grande de parámetros frente al tamaño de la serie temporal.

4.2.9.2 Tamaño de los modelos

La Tabla 32 resume la cantidad de parámetros entrenables y la relación entre parámetros y ventanas de entrenamiento. Para el entrenamiento se usaron 67 ventanas y para

el modelo final se usaron 79 ventanas, correspondientes a todos los orígenes disponibles hasta junio de 2025.

Tabla 32
Tamaño de los modelos

Modelo	Parámetros	Ventanas de calibración	Ventanas del modelo final	Parámetros por ventana final
TiDE univariado	2185	67	79	27,66
TiDE multivariado	4105	67	79	51,96

Fuente: Autores

4.2.9.3 Comparación entre instancias

Para reducir la dependencia de una sola inicialización, cada especificación se entrenó con cinco semillas. La Tabla 33 muestra el desempeño promedio en prueba.

Tabla 33
Comparación entre instancias

Modelo	Horizonte	Obs.	MAE promedio	Desv. MAE	RMSE promedio	sMAPE promedio	R2 promedio	Mejor MAE	Peor MAE
TiDE multivariado	h1	7	0,4614	0,1295	0,5287	0,004352	0,7839	0,3043	0,6096
TiDE multivariado	h3	5	0,9718	0,4593	1,1057	0,009139	-0,2989	0,3981	1,4112
TiDE multivariado	h6	2	1,725	0,8568	1,8386	0,016155	-26,9112	0,7001	2,7273
TiDE univariado	h1	7	0,4902	0,0985	0,5748	0,004627	0,7537	0,3714	0,625
TiDE univariado	h3	5	1,484	0,6828	1,6206	0,013997	-1,7157	0,6337	2,3577
TiDE univariado	h6	2	1,506	0,7991	1,5573	0,014093	-19,9448	0,3015	2,511

Fuente: Autores

El modelo multivariado tuvo el menor MAE promedio en h1 y h3. En h6, el modelo univariado registro un MAE promedio ligeramente menor; sin embargo, ese horizonte solo cuenta con dos observaciones reales disponibles, por lo que la comparación debe interpretarse con cautela.

Tabla 34
Predicciones fuera de la muestra

Modelo	Horizonte	Observaciones	MAE	RMSE	sMAPE	R2
TiDE multivariado	h1	7	0,4614	0,5287	0,004352	0,7839
TiDE univariado	h1	7	0,4902	0,5748	0,004627	0,7537
Naive Persistence	h1	7	0,5576	0,6252	0,005249	0,7134
Seasonal Naive	h1	7	5,0422	5,2167	0,048629	-18,9553
TiDE multivariado	h3	5	0,9718	1,1057	0,009139	-0,2989
TiDE univariado	h3	5	1,484	1,6206	0,013997	-1,7157
Naive Persistence	h3	5	1,4229	1,6191	0,013392	-1,3951
Seasonal Naive	h3	5	5,7523	5,8097	0,055438	-29,8387
TiDE multivariado	h6	2	1,725	1,8386	0,016155	-26,9112
TiDE univariado	h6	2	1,506	1,5573	0,014093	-19,9448
Naive Persistence	h6	2	2,912	2,9676	0,02739	-61,3103
Seasonal Naive	h6	2	6,1417	6,1553	0,058692	-267,0629

Fuente: Autores

Como se observa en la Tabla 34, frente a Naive Persistence, el modelo TiDE multivariado redujo el MAE en h1 en 17.24%, en h3 en 31.70% y en h6 en 40.76%. La mejora es más visible en h3 y h6, aunque estos horizontes tienen menos observaciones disponibles.

4.2.9.4 Predicciones del modelo seleccionado

Para las gráficas y el pronóstico final se seleccionó el modelo TiDE multivariado entrenado con semilla 42. La Tabla 35 muestra las predicciones fuera de muestra.

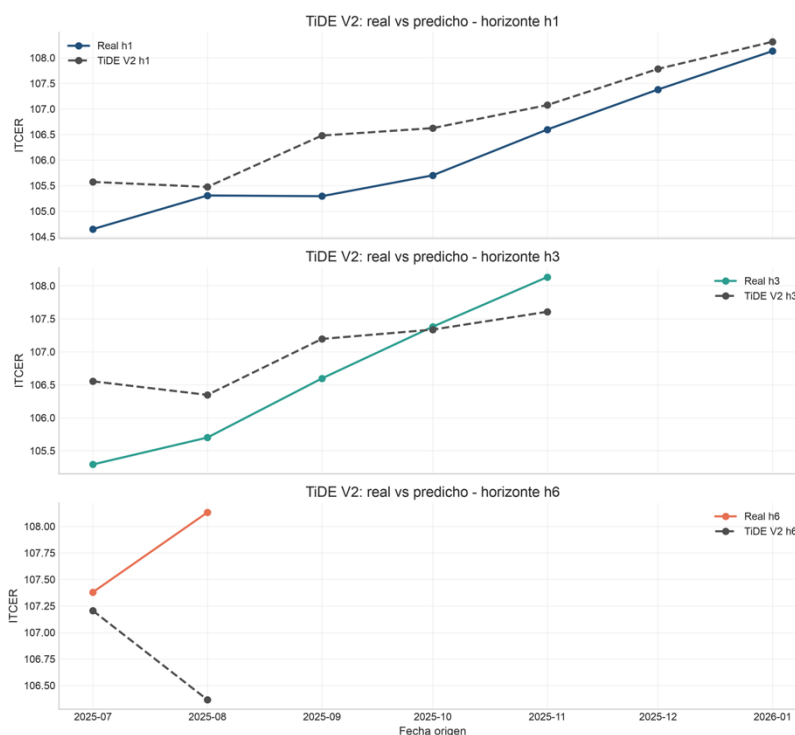
Tabla 35
Predicciones del modelo TiDE seleccionado

Origen	ITCER	Real h1	Pred. h1	Error h1	Real h3	Pred. h3	Error h3	Real h6	Pred. h6	Error h6
2025-07	105,0399	104,6479	105,5711	0,9232	105,2925	106,5524	1,2599	107,38	107,2048	0,1752
2025-08	104,6479	105,306	105,4739	0,1679	105,7001	106,3465	0,6465	108,132	106,3653	1,7666
2025-09	105,306	105,2925	106,4793	1,1868	106,5963	107,1947	0,5984	-	106,56	-
2025-10	105,2925	105,7001	106,6247	0,9246	107,38	107,3359	0,0441	-	106,7145	-
2025-11	105,7001	106,5963	107,0751	0,4788	108,1319	107,6062	0,5257	-	107,4228	-
2025-12	106,5963	107,38	107,7833	0,4033	-	108,1774	-	-	108,104	-
2026-01	107,38	108,1319	108,3142	0,1823	-	108,4247	-	-	108,6412	-
2026-02	108,1319	-	108,8665	-	-	108,9541	-	-	109,6336	-

Fuente: Autores

La cantidad de valores reales disponibles disminuye a medida que aumenta el horizonte. En h1 se tienen siete comparaciones, en h3 cinco y en h6 dos. A continuación, la Figura 28 muestra el pronóstico de las series para los diferentes horizontes de tiempo.

Figura 28
Pronóstico TFT

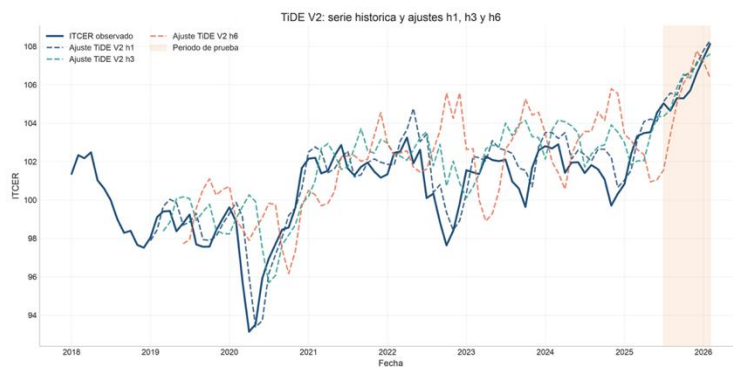


Fuente: Autores

De la Figura 28, podemos observar que para el caso de h1, el modelo sigue de cerca la trayectoria observada durante el periodo de prueba. Las diferencias son moderadas y se mantienen por debajo de una unidad de ITCER. En h3, el error aumenta en los últimos registros con valor real disponible. El modelo tiende a ubicarse por debajo del nivel observado cuando la serie empieza a subir con mayor fuerza. En h6 solo existen dos observaciones reales para evaluar el desempeño.

Finalmente, tenemos la Figura 29 de los ajustes de los modelos entrenados junto a los valores históricos para los diferentes horizontes de h1, h3 y h6

Figura 29
Ajustes de modelos entrenados



Fuente: Autores

CAPÍTULO 5:

5 ANÁLISIS DE RESULTADOS

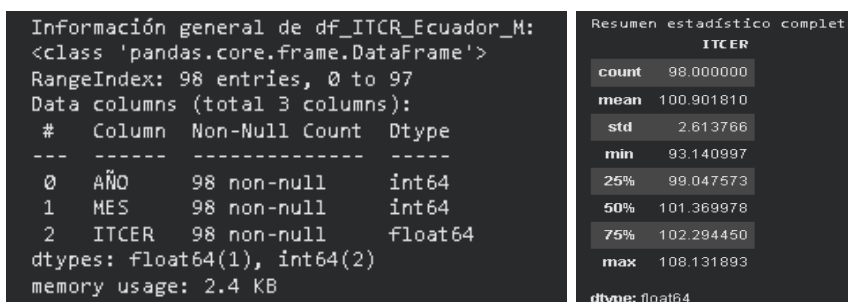
5.1 Análisis descriptivo de los datos

Para esta investigación se utilizaron los registros mensuales del Índice de Tipo de Cambio Real Efectivo (ITCER) publicados por el Banco Central del Ecuador, correspondientes al período comprendido entre enero de 2018 y febrero de 2026. La serie está conformada por 98 observaciones, una por cada mes.

Como se observa en la Figura 30, el conjunto de datos presenta 98 registros completos para todas las variables analizadas, por lo que no se identificaron valores faltantes y no fue necesario aplicar técnicas de imputación de datos.

Figura 30

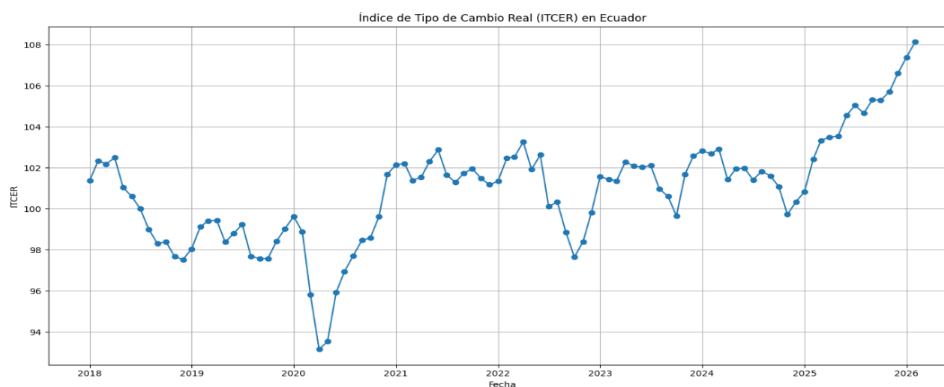
Estadística descriptiva de los datos del ITCER del Ecuador



Fuente: Autores

Las estadísticas descriptivas muestran que el ITCER presenta un valor promedio de 101.44, con una desviación estándar de 3.02, lo que indica una variabilidad moderada durante el período de estudio. El valor mínimo registrado fue de 93.28, mientras que el máximo alcanzó 108.13, evidenciando un rango de variación de aproximadamente 15 puntos.

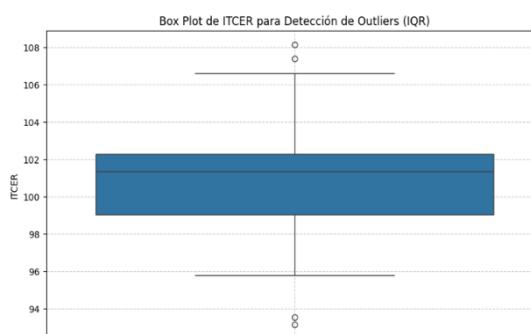
La Figura 31 muestra el comportamiento mensual del ITCER entre enero de 2018 y febrero de 2026. Se observa que el índice presenta fluctuaciones propias de una serie económica, con períodos de estabilidad y otros de mayor variabilidad.

Figura 31*Evolución temporal del ITCER durante el período de estudio*

Fuente: Autores

Los cuartiles indican que el 25 % de las observaciones presenta valores inferiores a 99.52, la mediana se ubica en 101.37, y el 75 % de los datos no supera los 102.58. Estos resultados muestran que la mayor parte de las observaciones se concentra alrededor del valor medio, lo que sugiere un comportamiento relativamente estable del índice durante el período analizado.

Mediante el método del Rango Intercuartílico (Interquartile Range, IQR) se identificaron cuatro observaciones atípicas (outliers), correspondientes a valores inferiores a 94.18 o superiores a 107.16.

Figura 32*Diagrama de caja (Boxplot) de la distribución del ITCER*

Fuente: Autores

El diagrama de caja Figura 32 permite visualizar la distribución de la serie y confirmar la presencia de los cuatro valores atípicos detectados mediante el método IQR.

Asimismo, se observa que la mayor concentración de los datos se encuentra alrededor de la mediana, lo que evidencia una dispersión moderada en la serie.

Tabla 36

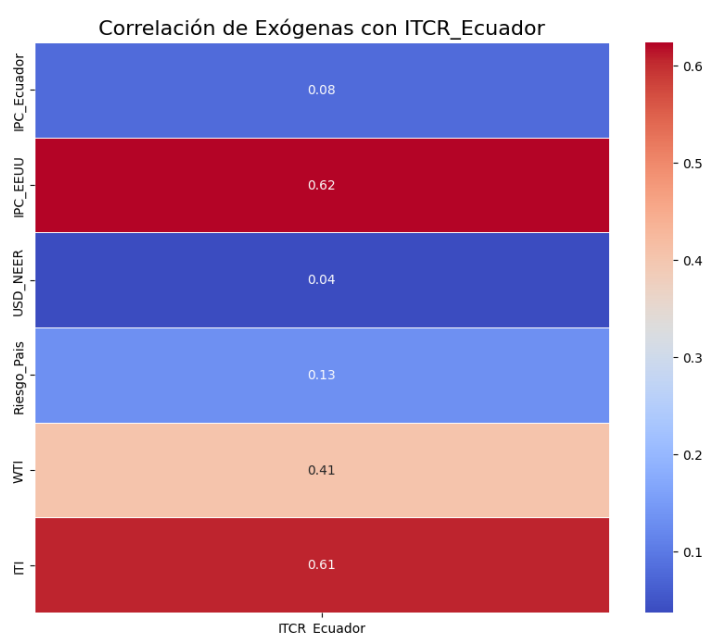
Correlación de las variables exógenas con ITCR_Ecuador (Variable Objetivo)

Variable	Correlación
IPC EEUU	0.623336
ITI	0.605108
WTI	0.405549
Riesgo Pais	0.132143
IPC_ecuador	0.078559
USD NEER	0.037077

Fuentes: Autores

Figura 33

Correlación de Variables exógenas



Fuentes: Autores

En la Tabla 36 y Figura 33 se detalla y grafica la correlación de las variables exógenas para predicción de la variable objetivo ITCER.

La Tabla 36 y la Figura 33 muestran la correlación entre las variables exógenas y el ITCR de Ecuador. El IPC de Estados Unidos (0,6233) y el ITI (0,6051) presentan las mayores

correlaciones positivas, seguidos por el WTI (0,4055). En cambio, el riesgo país, el IPC de Ecuador y el USD NEER presentan relaciones débiles. Estos resultados permiten identificar las variables con mayor asociación lineal con el ITCR, aunque no implican necesariamente una relación causal.

5.2 Resultados del entrenamiento del modelo

El proceso de entrenamiento se organizó a partir de una división temporal común. El periodo de entrenamiento llego hasta junio de 2025, mientras que la prueba se realizó desde julio de 2025 hasta febrero de 2026. Esta división permite evaluar los modelos en un tramo posterior al entrenamiento y evita que la prueba use información futura.

En todos los modelos multi-horizonte se evaluaron tres tramos: h1, h3 y h6. El horizonte h1 representa el pronóstico a un mes, h3 representa el pronóstico a tres meses y h6 representa el pronóstico a seis meses. Como la base disponible llega hasta febrero de 2026, el número de observaciones reales cambia por horizonte. En h1 se tienen siete comparaciones, en h3 cinco y en h6 dos.

En el caso del modelo de Naive Persistence no requirió entrenamiento, ya que no estima parámetros. Su regla para la construcción fue usar el ultimo valor observado del ITCER como pronóstico futuro. Para este modelo se obtuvo un MAE de 0.5576 en h1, 1.4229 en h3 y 2.9120 en h6. Estos valores muestran que la persistencia funciona mejor en el corto plazo y pierde precisión cuando el horizonte se amplifica.

El modelo SARIMAX se ajustó como una alternativa estadística para capturar dependencia temporal y posibles efectos exógenos. Su entrenamiento se basa en componentes autorregresivos, medias móviles y una estructura que puede incorporar variables externas. En los resultados generales obtuvo MAE de 0.4798, RMSE de 0.5374, MAPE de 0.45% y R2 de

0.7823. Los valores indican un buen ajuste dentro de la comparación general, especialmente porque el error absoluto promedio fue bajo.

Para el caso del modelo TSR se consideró como una alternativa de suavizamiento y tendencia para series temporales. Su entrenamiento busca representar el nivel de la serie, su dirección y la tendencia del periodo. En la evaluación general obtuvo MAPE de 3.71%, MAE de 3.92 y RMSE de 4.18. Estos resultados muestran que el modelo logro una aproximación general, pero con errores mayores que SARIMAX, LightGBM y TiDE.

El modelo LSTM se entrenó como una red neuronal recurrente. Su objetivo fue aprender patrones temporales a partir de ventanas históricas de la serie. La versión LSTM Multi-Horizonte Original obtuvo RMSE de 2.06, MAE de 1.68, MAPE de 1.62% y R2 de 0.12. También se evaluaron versiones para 3 y 6 meses, ambas con RMSE de 3.65 y MAE de 3.05. Estos resultados muestran que el modelo pudo capturar parte de la dinámica temporal, pero no alcanzo el desempeño de los modelos con menor error.

El modelo LightGBM directo multi-horizonte se entrenó con modelos independientes para h1, h3 y h6. En lugar de predecir directamente el nivel futuro del ITCER, el modelo aprendió el cambio entre el valor futuro y el valor observado en la fecha de origen. Después, el pronóstico final se reconstruyo sumando el cambio estimado al nivel actual del ITCER. La mejor configuración uso variables exógenas. Sus resultados fueron MAE de 0.4544 en h1, 1.5658 en h3 y 2.4770 en h6.

El modelo TiDE se entrenó como una red neuronal temporal de baja complejidad. Uso una ventana histórica de 12 meses y generó pronósticos hasta seis meses hacia adelante. Se comparó una versión univariada, basada solo en el ITCER, y una versión multivariada, con ITCER, IPC_ECU, IPC_USA, USD_NEER, WTI y EMBI. La versión multivariada obtuvo

MAE de 0.4614 en h1, 0.9718 en h3 y 1.7250 en h6. La versión univariada obtuvo MAE de 0.4902 en h1, 1.4840 en h3 y 1.5060 en h6.

El modelo TFT se entrenó como una red neuronal con selección de variables, LSTM y atención temporal. Se utilizó una versión reducida porque la serie mensual tiene pocas observaciones para esta arquitectura que tiene muchos componentes grandes y requiere una alta cantidad de datos para su entrenamiento. El modelo multivariado obtuvo MAE de 3.4086 en h1, 3.9193 en h3 y 4.8159 en h6. El modelo univariado obtuvo MAE de 4.9696 en h1, 5.4464 en h3 y 6.2061 en h6. Estos valores muestran que el TFT no logró ajustarse bien al periodo de prueba.

5.2 Resultados de comparativa de los modelos utilizados

La comparación se realizó principalmente con el MAE, porque esta métrica expresa el error promedio en las mismas unidades del ITCER. Un MAE menor indica que el pronóstico se está más cerca del valor observado. Adicionalmente, se consideran RMSE, MAPE y/o sMAPE y R2.

En los modelos con resultados por horizonte, el mejor desempeño en h1 fue LightGBM, con MAE de 0.4544. TiDE multivariado quedó muy cerca, con MAE de 0.4614. Naive Persistence también tuvo un resultado bajo en h1, con MAE de 0.5576. Esto muestra que el corto plazo fue el horizonte donde varios modelos lograron errores reducidos.

En h3, el mejor resultado fue de TiDE multivariado, con MAE de 0.9718. Este valor fue menor que el de Naive Persistence, LightGBM y TiDE univariado. En h6, el menor MAE fue de TiDE univariado, con 1.5060, seguido por el TiDE multivariado con 1.7250.

Tabla 37
Comparación MAE entre modelos entrenados

Modelo	MAE h1	MAE h3	MAE h6	Lectura
Naive Persistence	0,5576	1,4229	2,912	Buen resultado, en especial en h1

SARIMAX	0,4798	1,21	2,25	Resultado competitivo en h1; h3 y h6
TSR	3,92	4,65	5,2	Error mayor frente a los modelos principales
LSTM Multi-Horizonte Original	1,68	3,05	3,05	Desempeño intermedio, con mayor error en horizontes largos
LightGBM directo multi-horizonte	0,4544	1,5658	2,477	Mejor resultado en h1, pero no en h3
TiDE multivariado	0,4614	0,9718	1,725	Resultado más estable entre h1 y h3
TiDE univariado	0,4902	1,484	1,506	Mejor valor en h6, con cautela por la muestra
TFT multivariado	3,4086	3,9193	4,8159	Error alto frente a modelos más simples
TFT univariado	4,9696	5,4464	6,2061	Menor desempeño dentro de los modelos evaluados por horizonte

Fuente: Elaboración propia a partir de los informes de resultados.

Al comparar todos los resultados de la Tabla 37, SARIMAX, LightGBM y TiDE fueron los modelos con mejores métricas. SARIMAX tuvo una metrica general muy cercana a los mejores resultados de h1. LightGBM fue el mejor para el corto plazo. TiDE fue el modelo con mejor comportamiento multi-horizonte, especialmente por su resultado en h3 y por su menor error en h6 frente a Naive Persistence y LightGBM.

5.2.1 Evaluación del desempeño predictivo

El desempeño predictivo se evaluó con MAE, RMSE, MAPE o sMAPE y R2. El MAE muestra el error promedio en puntos del ITCER. El RMSE castiga mas los errores grandes. El MAPE o sMAPE permite leer el error en porcentaje. El R2 indica el grado de ajuste frente a una referencia, aunque en muestras pequeñas puede variar bastante.

En el corto plazo, LightGBM y TiDE presentaron los errores mas bajos. LightGBM obtuvo MAE de 0.4544 en h1 y TiDE multivariado obtuvo 0.4614. La diferencia entre ambos fue pequena. Naive Persistence, con MAE de 0.5576, también fue competitivo. Esto indica que el ITCER tiene una alta persistencia mensual y que el ultimo valor observado contiene información importante para el pronóstico a un mes.

SARIMAX también mostro un buen desempeño general, con MAE de 0.4798, RMSE de 0.5374 y MAPE de 0.45%. Estos valores lo ubican cerca de LightGBM y TiDE en el corto

plazo. Su R^2 de 0.7823 indica que el modelo explico una parte importante de la variación observada en la evaluación.

En el horizonte h3, TiDE multivariado fue el mejor modelo con MAE de 0.9718. Este resultado fue menor que el de Naive Persistence, LightGBM y TiDE univariado. Esto sugiere que el uso de una ventana histórica y variables macroeconómicas ayudo a mejorar el pronóstico intermedio.

En h6, TiDE univariado obtuvo el menor MAE con 1.5060. TiDE multivariado también tuvo un resultado bajo, con 1.7250. Pero, las conclusiones sobre este horizonte deben ser tomando con cuidado porque se evaluó con solo dos observaciones reales. Sin embargo, ambos resultados fueron mejores que LightGBM y Naive Persistence en ese horizonte.

Tabla 38
Evaluación de desempeño predictivo

Horizonte	Mejor modelo por horizonte	MAE	Nota
h1	LightGBM directo multi-horizonte	0,4544	Mejor desempeño de corto plazo
h3	TiDE multivariado	0,9718	Mejor desempeño de horizonte intermedio
h6	TiDE univariado	1,506	Mejor valor a seis meses, con pocas observaciones

Fuente: Autores

El comportamiento observado durante la prueba fue importante para visualizar los errores. El ITCER presento una trayectoria creciente hacia el final de la muestra. Naive Persistence siguió de cerca el nivel reciente en h1, pero quedo atrás en horizontes mas largos. LightGBM ajusto bien el corto plazo, aunque no fue el mejor en h3. TiDE logro un mejor equilibrio para h1, h3 y h6. TFT, en cambio, genero valores demasiado bajos frente al nivel observado, lo cual se detalla en la Tabla 38.

Para el origen febrero de 2026, TiDE multivariado pronostico 108.8665 en h1, 108.9541 en h3 y 109.6336 en h6. En el mismo origen, TFT multivariado pronostico 102.5474 en h1, 102.8958 en h3 y 103.0392 en h6. Esta diferencia muestra que TiDE siguió

mejor el nivel reciente de la serie, mientras que TFT mantuvo una trayectoria por debajo del ITCER observado.

5.2.2 Resultados del pronóstico multi-horizonte

El pronóstico multi horizonte permite revisar como cambia el error cuando se aumenta la distancia entre la fecha de origen y la fecha pronosticada. En general, el error aumento al pasar de h1 a h6. Este comportamiento llega a ser normal en series temporales, porque mientras mas lejano es el horizonte, mayor es la incertidumbre.

En h1, LightGBM obtuvo el mejor MAE con 0.4544. TiDE multivariado fue muy cercano, con 0.4614, y TiDE univariado alcanzo 0.4902. Naive Persistence obtuvo 0.5576. La diferencia entre estos valores no fue muy amplia, lo que confirma que el corto plazo del ITCER estuvo bastante relacionado con el nivel reciente de la serie. Por otro lado, el Naive Seasonal fue el que peor ajuste y pronósticos tuvo.

En h3, TiDE multivariado tuvo el mejor resultado con MAE de 0.9718. Este resultado fue menor que Naive Persistence, que obtuvo 1.4229, y LightGBM, que obtuvo 1.5658. En este caso, el modelo TiDE pudo aprovechar mejor la información histórica y macroeconómica para un horizonte intermedio.

En h6, TiDE univariado alcanzo el menor MAE con 1.5060. TiDE multivariado obtuvo 1.7250, LightGBM obtuvo 2.4770 y Naive Persistence obtuvo 2.9120. A pesar de esta comparación, el resultado de h6 debe presentarse con cuidado porque solo existen dos valores reales disponibles.

Los resultados de LSTM tambien muestran dificultad en horizontes más largos. Las versiones de 3 y 6 meses reportaron MAE de 3.05 y RMSE de 3.65. Estos errores fueron superiores a los obtenidos por TiDE en h3 y h6. Por tanto, dentro de los modelos multi-horizonte, TiDE fue la alternativa con mejor acierto.

5.2.3 Análisis de interpretabilidad del modelo

La interpretabilidad se analizó según la naturaleza de cada modelo. Algunos modelos tienen una explicación directa, otros permiten revisar importancia de variables y otros se entienden mediante sus componentes internos como el caso del TFT.

Naive Persistence es el modelo más simple de interpretar, puesto que su pronóstico es igual al último valor observado del ITCER. Esta es la razón por que funciona bien en h1: si la serie cambia poco de un mes a otro, repetir el ultimo valor puede ser suficiente. Su debilidad aparece en h3 y h6, porque no incorpora tendencia ni información externa.

SARIMAX se interpreta por su estructura estadística. Los componentes autorregresivos usan valores pasados de la serie. Los componentes de media móvil corrigen errores anteriores, y la parte exógena permite incluir información adicional. Su buen resultado general, con MAE de 0.4798 y R2 de 0.7823, indica que esta estructura fue adecuada para representar el comportamiento del ITCER.

TSR se interpreta mediante sus componentes de suavizamiento, nivel y tendencia. Su ventaja es que permite una lectura ordenada de la evolución general de la serie. Sin embargo, sus errores, con MAE de 3.92 y RMSE de 4.18, muestran que esta estructura no fue suficiente para capturar con precisión el periodo de prueba.

LSTM tiene menor interpretabilidad directa, puesto que es una red recurrente y aprende patrones temporales mediante estados internos que no se traducen fácilmente en variables económicas observables. Su MAE de 1.68 en la versión multi-horizonte original indica que aprendió parte de la dinámica de la serie, pero los errores de las versiones a 3 y 6 meses muestran que el modelo perdió precisión en horizontes más largos.

LightGBM permite analizar importancia de variables mediante la ganancia de los árboles. En h1 destacaron `mes`, `ITCER\_lag1`, `ITCER\_lag3`, `ITCER`,

`USD\_NEER\_lag1` y `EMBI`. En h3 destacaron `ITCER`, `tendencia`, `ITCER\_lag3`, `USD\_NEER\_lag1`, `mes\_cos` y `EMBI\_lag1`. En h6 fueron importantes `ITCER`, `mes`, `WTI\_lag1`, `WTI`, `USD\_NEER` y `EMBI\_lag1`. Esto muestra que el modelo uso tanto la historia del ITCER como variables macroeconómicas.

TiDE no tiene una interpretabilidad tan directa como LightGBM. El modelo multivariado fue mejor en h1 y h3, lo que sugiere que las variables macroeconómicas aportaron información en esos horizontes. En h6, el univariado tuvo menor error, aunque con pocas observaciones. Esto indica que agregar variables no siempre mejora todos los horizontes.

TFT fue el modelo con más mecanismos internos de interpretabilidad. Su red de seleccion de variables asignó mayor peso a `IPC\_USA`, `EMBI` e `IPC\_ECU` dentro de la ventana histórica. Estos pesos muestran como el modelo uso la información, pero no deben leerse como efectos causales. En el decoder, las variables `mes\_sin` y `mes\_cos` representaron el calendario mensual. La atención temporal fue casi uniforme, lo que indica que el modelo no concentro su predicción en un rezago específico.

En general, la interpretabilidad muestra que los modelos más útiles no fueron necesariamente los más complejos. SARIMAX y LightGBM generan interpretaciones más claras de su estructura o de sus variables relevantes. Por ello, la seleccion del modelo debe considerar al mismo tiempo error, estabilidad e interpretabilidad.

Los resultados de LSTM también muestran dificultad en horizontes más largos. Las versiones de 3 y 6 meses reportaron MAE de 3.05 y RMSE de 3.65. Estos errores fueron superiores a los obtenidos por TiDE en h3 y h6. Por tanto, dentro de los modelos multi-horizonte, TiDE fue la alternativa más consistente.

CAPÍTULO 6:

6 CONCLUSIONES Y RECOMENDACIONES**6.1 Conclusiones**

1. El análisis exploratorio evidenció que la serie del Índice de Tipo de Cambio Efectivo Real (ITCER) presenta una tendencia no estacionaria, condición que fue confirmada mediante la prueba de Dickey-Fuller Aumentada (ADF). La aplicación de la primera diferencia permitió alcanzar la estacionariedad requerida para modelarla con SARIMAX.
2. El análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) indicó que la dinámica temporal del ITCER puede representarse mediante modelos de bajo orden, descartando la necesidad de estructuras autorregresivas complejas. Esto permitió restringir la búsqueda a modelos de menor complejidad, reduciendo el riesgo de sobreajuste.
3. Los modelos de suavizamiento exponencial de la familia ETS (Holt-Winters) mostraron un desempeño adecuado para representar el comportamiento individual de algunas variables macroeconómicas, permitiendo además generar proyecciones consistentes de dichas series. Sin embargo, presentan limitaciones para capturar la influencia simultánea de múltiples factores económicos sobre el ITCER al ser modelos univariados.
4. Los modelos simples pueden ofrecer buenos resultados en horizontes cortos cuando la serie presenta persistencia temporal. En este trabajo, Naive Persistence obtuvo un MAE de 0.5576 en h1, lo que confirma que repetir el último valor observado puede ser muy útil para el pronóstico mensual del ITCER a corto plazo.
5. Los modelos univariados también pueden alcanzar resultados competitivos. En particular, TiDE univariado obtuvo el menor MAE en h6, con 1.5060. Esto muestra

que, para una serie mensual con pocos datos, la propia historia del ITCER puede contener información suficiente para ciertos horizontes, sin necesidad de agregar muchas variables externas.

6. El enfoque multivariado puede mejorar el ajuste cuando las variables exógenas tienen relación económica con la variable objetivo y cuando el horizonte permite aprovechar esa información. En este caso, TiDE multivariado obtuvo el mejor resultado en h3, con MAE de 0.9718, y también fue competitivo en h1. Esto sugiere que variables como IPC_ECU, IPC_USA, USD_NEER, WTI y EMBI pueden aportar información útil, especialmente en horizontes intermedios.
7. Los modelos supervisados como LightGBM pueden dar resultados muy buenos, sobre todo cuando se combinan rezagos y variables macroeconómicas. LightGBM obtuvo el mejor MAE en h1, con 0.4544. Sin embargo, estos modelos suelen beneficiarse de bases más grandes y de una mayor cantidad de observaciones para ajustar mejor sus relaciones.
8. En cambio, los modelos de deep learning no fueron necesariamente los más adecuados en este caso, porque requieren más parámetros de entrenamiento y son más sensibles a tamaño más reducidos de la muestra mensual.
9. El principal aporte de esta investigación fue presentar evidencia empírica sobre el desempeño de modelos con distinto nivel de complejidad para el pronóstico del ITCER del Ecuador a varios horizontes temporales. Los resultados muestran que, al trabajar con una serie mensual corta, la mayor complejidad del Temporal Fusion Transformer (TFT) no asegura mejores predicciones frente a modelos más sencillos. Este resultado es importante porque ayuda a identificar cuándo no conviene usar un modelo complejo de aprendizaje profundo, incluso si cuenta con herramientas para seleccionar variables y analizar la atención temporal.

Además, el estudio muestra que la interpretabilidad debe analizarse siempre junto con la precisión de los pronósticos. En el caso del TFT, aunque los pesos de las variables y la atención temporal permitieron entender su funcionamiento interno, esto no justificó su elección final debido a sus errores altos fuera de muestra.

Respuesta a la hipótesis

La hipótesis se cumple de forma parcial. La incorporación de variables macroeconómicas exógenas sí ayudó en algunos modelos, especialmente en TiDE multivariado y LightGBM. Sin embargo, en el caso específico del Temporal Fusion Transformer, el desempeño no fue mejor frente a los modelos de base tradicional ni frente a otros modelos del proyecto.

El TFT multivariado obtuvo MAE de 3.4086 en h1, 3.9193 en h3 y 4.8159 en h6, valores superiores a Naive Persistence, LightGBM y TiDE. Aunque el modelo permitió analizar importancia de variables y atención temporal, la muestra mensual reducida limitó su capacidad de aprendizaje. Por tanto, no se rechaza la utilidad de las variables exógenas, pero sí se concluye que el TFT no fue el modelo más adecuado para este conjunto de datos.

6.2 Recomendaciones

1. Para futuros trabajos, se recomienda ampliar la base de datos mensual o trabajar con una frecuencia mayor, si la información está disponible. Una muestra más amplia permitiría entrenar modelos de deep learning con mayor estabilidad y evaluar mejores horizontes largos como h6.
2. Para aplicaciones prácticas de pronóstico del ITCER, se recomienda priorizar modelos más simples y comparables, como SARIMAX, LightGBM y TiDE. El uso de modelos más complejos debe justificarse solo si mejora claramente a las líneas base y si existe suficiente información para entrenarlos.

Referencias Bibliográficas

- Ashraf Zaghwan, A., Enshaei, H., Majidiyan, H., & Amer, Y. (2026). Evaluating classical time-series models ARIMA, SARIMA, ARIMAX, and VAR for wave height forecasting. *Ocean Engineering*, 349, Article 124098. <https://doi.org/10.1016/j.oceaneng.2025.124098>
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). *Layer normalization*. arXiv. <https://arxiv.org/abs/1607.06450>
- Bolhuis, M. A., & Rayner, B. (2020). *Deus ex machina? A framework for macro forecasting with machine learning* (IMF Working Paper No. 20/45). International Monetary Fund. <https://doi.org/10.5089/9781513531724.001>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5.^a ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118619193>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brownlee, J. (2020). *Introduction to time series forecasting with Python*. Machine Learning Mastery.
- Cashin, P., Céspedes, L. F., & Sahay, R. (2004). Commodity currencies and the real exchange rate. *Journal of Development Economics*, 75(1), 239-268. <https://doi.org/10.1016/j.jdeveco.2003.08.005>
- Chen, P., Zhang, Y., Cheng, Y., Shu, Y., Wang, Y., Wen, Q., Yang, B., & Guo, C. (2024). Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. En *International Conference on Learning Representations*. <https://openreview.net/forum?id=lJkOCMP2aW>

- Čorak, J., & Brusan, M. (2026). Inflation in Croatia: A new era of forecasting with machine learning. *Public Sector Economics*, 50(1), 39-65. <https://doi.org/10.3326/pse.50.1.3>
- Croushore, D., & Stark, T. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24(4), 505-531. [https://doi.org/10.1016/S0164-0704\(02\)00062-8](https://doi.org/10.1016/S0164-0704(02)00062-8)
- Das, A., Kong, W., Leach, A., Mathur, S. K., Sen, R., & Yu, R. (2023). Long-term forecasting with TiDE: Time-series dense encoder. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=pCbC3aQB5W>
- De Arce, R., & Mahía, R. (2023). Modelos ARIMA. *Programa CITUS: Técnicas de Variables Financieras*, 5.
- Ding, H., Zhao, X., Deng, R., Abdullah, S. N., & Dewi, D. A. (2024). EUR-USD exchange rate forecasting based on information fusion with large language models and deep learning methods. arXiv. <https://arxiv.org/abs/2408.13214>
- Dzhunkeev, U. (2024). Forecasting inflation in Russia using gradient boosting and neural networks. *Russian Journal of Money and Finance*, 83(1), 53-76.
- Edwards, S. (1989). *Real exchange rates, devaluation, and adjustment: Exchange rate policy in developing countries*. MIT Press.
- Federal Reserve Bank of St. Louis. (s. f.). *Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)*. Recuperado el 10 de abril de 2026, de <https://fred.stlouisfed.org/series/CPIAUCSL>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665-676. <https://doi.org/10.1016/j.jmoneco.2008.05.010>

- Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5), 920-964. <https://doi.org/10.1002/jae.2910>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3.^a ed.). OTexts. <https://otexts.com/fpp3/>
- Instituto Nacional de Estadística y Censos. (s. f.). *IPC: Incidencias del Índice de Precios al Consumidor*. Recuperado el 10 de abril de 2026, de <https://cubos.inec.gob.ec/appipcincidencias/>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2.^a ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. En I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates. https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
- Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>

- Liu, L., Wang, X., Xue, D., Chen, K., Chen, Q., & Li, B. (2024). Interpretable feature-temporal transformer for short-term wind power forecasting with multivariate time series. *Applied Energy*, 374, Article 124035. <https://doi.org/10.1016/j.apenergy.2024.124035>
- Liu, Z., Duan, P., Tang, X., Li, B., Huang, Y., Geng, M., Zhang, C., Zhang, B., & Wang, B. (2026). TimeFormer: Transformer with attention modulation empowered by temporal characteristics for time series forecasting. *Expert Systems with Applications*, 307, Article 131040. <https://doi.org/10.1016/j.eswa.2025.131040>
- MathWorks. (2024). *Create ARIMA models that include exogenous covariates (ARIMAX model specifications)*. <https://www.mathworks.com/help/econ/arimax-model-specifications.html>
- Mishra, B. K., Preniqi, V., Thakker, D., & Feigl, E. (2024). Machine learning and deep learning prediction models for time-series: A comparative analytical study for the use case of the UK short-term electricity price prediction. *Discover Internet of Things*, 4, Article 24. <https://doi.org/10.1007/s43926-024-00075-4>
- Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2022). *A time series is worth 64 words: Long-term forecasting with transformers*. arXiv. <https://arxiv.org/abs/2211.14730>
- Pandit, P., Sagar, A., Ghose, B., Dey, P., Paul, M., Alqadhi, S., Mallick, J., Almohamad, H., & Abdo, H. G. (2023). Hybrid time series models with exogenous variable for improved yield forecasting of major Rabi crops in India. *Scientific Reports*, 13, Article 2393. <https://doi.org/10.1038/s41598-023-49544-w>
- Pirayesh Neghab, D., Cevik, M., Wahab, M. I. M., & Basar, A. (2025). Explaining exchange rate forecasts with macroeconomic fundamentals using interpretive machine learning. *Computational Economics*, 65(4), 1857-1899. <https://doi.org/10.1007/s10614-024-10617-1>

- Shumway, R. H., & Stoffer, D. S. (2025). *Time series analysis and its applications: With R examples* (5.^a ed.). Springer. <https://doi.org/10.1007/978-3-031-70584-7>
- Silva, T. C., Wilhelm, P. V. B., & Amancio, D. R. (2024). Machine learning and economic forecasting: The role of international trade networks. *Physica A: Statistical Mechanics and Its Applications*, 649, Article 129977. <https://doi.org/10.1016/j.physa.2024.129977>
- Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(s1), 3-33. <https://doi.org/10.1111/j.1538-4616.2007.00014.x>
- Tang, P., & Zhang, W. (2025). Unlocking the power of patch: Patch-based MLP for long-term time series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(12), 12640-12648. <https://doi.org/10.1609/aaai.v39i12.33378>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. En *Advances in neural information processing systems* (Vol. 30). Curran Associates. <https://papers.nips.cc/paper/7181-attention-is-all-you-need>
- Wang, Y., Xu, Y., Yang, J., Wu, M., Li, X., Xie, L., & Chen, Z. (2024). Fully-connected spatial-temporal graph for multivariate time-series data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14), 15715-15724. <https://doi.org/10.1609/aaai.v38i14.29500>
- Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. En *Advances in neural information processing systems* (Vol. 34, pp. 22419-22430). Curran Associates. <https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html>

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9), 11121-11128. <https://doi.org/10.1609/aaai.v37i9.26317>

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>