



Universidad del
Rosario

| Escuela de Ingeniería,
Ciencia y Tecnología

**PRONOSTICANDO EL VOLUMEN DEL MERCADO INTERBANCARIO DE
DIVISAS: CASO COLOMBIANO**

MAGISTER EN MATEMÁTICAS APLICAS Y CIENCIAS DE LA COMPUTACIÓN

Paula Andrea Torres Medina

Dirección:

Gabriel Camilo Pérez

Universidad del Rosario

Escuela de Ingeniería, Ciencia y Tecnología

Maestría en matemáticas aplicadas y ciencias de la computación

2023

Índice

1. Planteamiento del problema y justificación	7
2. Objetivos	7
2.1. Objetivo General	7
2.2. Objetivos específicos	8
3. Marco Teórico	8
3.1. Mercado FOREX	8
3.2. Modelos Estadísticos	9
3.2.1. Modelos de series de tiempo	9
3.3. Modelos de Aprendizaje Automático	12
3.3.1. Memoria de Largo y Corto Plazo (LSTM)	13
3.3.2. Temporal Convolutional Networks (TCN)	15
3.3.3. Extreme Gradient Boosting (XGBOOST)	16
3.4. Métricas de evaluación	19
4. Estado del arte	19
5. Metodología	21
5.1. Recolección de datos	21
5.2. Preprocesamiento de datos	22
5.3. Validación cruzada	24
5.4. Implementación de los modelos	25
5.4.1. Modelo Autoregresivo	26
5.4.2. Modelo LSTM	29
5.4.3. Modelo TCN	30
5.4.4. Modelo XGBOOST	30
5.5. Optimización con Optuna	31
6. Resultados	31
7. Discusión	36

8. Conclusiones	37
9. Anexos	43
9.1. Modelos autoregresivos univariados	43
9.2. Modelos autoregresivos con variables auxiliares	45
9.3. Optimización	48
9.4. Test Diebold-Mariano	48

Lista de tablas

1.	Lista de variables	21
2.	Pruebas raíz unitaria	27
3.	Hiperparámetros LSTM	29
4.	Métricas por modelo	32
5.	Coeficientes ARMA(2,2) - volumen suavizado 5 días	43
6.	Coeficientes AR(5) - Δ volumen suavizado 40 días	43
7.	Coeficientes modelos para el Monto negociado	44
8.	Espacio de búsqueda hiperparámetros	48
9.	Test Diebold-Mariano volumen transacciones 5 días	49
10.	Test Diebold-Mariano monto negociado 5 días	49
11.	Test Diebold-Mariano volumen transacciones 40 días	50
12.	Test Diebold-Mariano monto negociado 40 días	50

Lista de figuras

1.	Estructura básica de una red LSTM Fuente: [1]	13
2.	Detalle de una capa de LSTM Fuente: [1]	14
3.	Cómo se ve la dilatación Fuente: [2]	16
4.	Estructura general de una TCN Fuente: [2]	17
5.	Estructura XGBOOST Elaboración propia	18
6.	Volumen transacciones - Suavizado	24
7.	Monto negociado - Suavizado	25
8.	Ejemplo Validación cruzada para series temporales	26
9.	Matrices de correlación	28
10.	Pronósticos series suavizadas 5 días	34
11.	Pronósticos series suavizadas 40 días	35
12.	Ajuste modelo ARMA volumen suavizado 5 días	43
13.	Ajuste modelo AR Δ volumen suavizado 40 días	44
14.	Ajuste modelo ARMA monto suavizado 5 días	44
15.	Ajuste modelo ARMA Δ monto suavizado 40 días	45
16.	Ajuste modelo ARMAX volumen suavizado 5 días	46
17.	Ajuste modelo ARMAX monto suavizado 5 días	46
18.	Ajuste modelo ARX Δ volumen suavizado 40 días	47
19.	Ajuste modelo ARMAX Δ monto suavizado 40 días	47

Agradecimientos

Deseo expresar mi sincero agradecimiento a mi mamá y a mi familia por brindarme un constante apoyo, comprensión y aliento a lo largo de este desafiante proceso. Sus palabras de aliento fueron un gran respaldo en los momentos más difíciles, otorgándome la fuerza necesaria para perseverar y seguir adelante.

Asimismo, deseo extender mi gratitud a mi compañero y amigo, Tomás Montealegre, por su invaluable apoyo y sus valiosos comentarios. Su apoyo incondicional ha sido fundamental en el desarrollo de este trabajo.

RESUMEN

En este trabajo se estudian las fortalezas y debilidades de los modelos de pronóstico del volumen de transacciones del mercado colombiano interbancario de divisas, generado por un modelo basado en árboles de decisión y dos tipos de redes neuronales, las Long short term memory y las temporal convolutional nexworks, comparados con los modelos econométricos tradicionales para el estudio de series de tiempo.

Palabras clave: Mercado FOREX, Análisis de series de tiempo, XGBOOST, Red LSTM, Red TCN

ABSTRACT

This paper studies the strengths and weaknesses of forecast models of the volume of transactions in the Colombian forex market. It analyzes a model based on decision trees and two types of neural networks, namely Long Short-Term Memory (LSTM) and Temporal Convolutional Networks (TCN), comparing them with traditional econometric models for the study of time series.

Keywords: FOREX Market, Time series analysis, XGBOOST, LSTM network, TCN Network

1. Planteamiento del problema y justificación

En una era donde la economía global está tan interconectada, el intercambio de bienes y servicios entre países toma un lugar relevante dentro de la economía, esto se evidencia en el creciente volumen de transacciones que presenta el mercado interbancario de divisas. Para abril de este año, el Banco de Pagos Internacionales¹ reportó en su encuesta trianual que diariamente se movían \$7.5 billones de dólares, lo que representa un aumento del 14 % respecto a 2019 [3].

Dada su importancia y volatilidad, este mercado ha sido objeto de diversos estudios, que además de caracterizarlo, buscan pronosticar su comportamiento. A pesar, de que el enfoque más usual apunta al pronóstico de las tasas de cambio, y se han hecho varios estudios al respecto [4] [5] [6], el volumen de transacciones no ha sido tan explorado. Algunos ejemplos son [7], [8], donde se han aplicado desde modelos de series de tiempo heterocedásticos hasta técnicas de minería de texto.

Así pues, el análisis y pronóstico del comportamiento del mercado cambiario de divisas colombiano es necesario y muy útil para los agentes involucrados con operaciones de comercio exterior, así como para inversores y entidades financieras que prestan servicios relacionados, porque permite una toma de decisiones informada para llevar a cabo compras o ventas de divisas de tal forma que se optimicen las operaciones y sus objetivos de negocio [9]. Lograr eso, teniendo en cuenta la situación política y económica actual es de suma importancia y gran diferencial para los inversores (empresas o personas naturales) y para los países donde operan.

2. Objetivos

2.1. Objetivo General

El objetivo de este trabajo es determinar, de un grupo de técnicas seleccionado, cuál presenta el mejor desempeño para pronosticar el volumen de las transacciones de USD-COP que se realizan en el mercado interbancario de divisas colombiano.

¹<https://www.bis.org/>

2.2. Objetivos específicos

- Establecer una medida del nivel de certeza que producen los distintos pronósticos con el fin de poder compararlos.
- Comparar los resultados de los diferentes modelos para elegir el mejor. Más certero y eficaz a la hora de pronosticar.
- Identificar las fortalezas y debilidades de cada modelo cuando son aplicados en el contexto del mercado cambiario USD/COP.

3. Marco Teórico

En esta sección, se busca exponer los conceptos económicos relacionados con el mercado cambiario de divisas, además de los modelos estadísticos y de aprendizaje de máquina que se emplearán en el desarrollo de este trabajo. Estos modelos se presentarán de forma progresiva, desde los más simples hasta los más complejos. Comenzando con los modelos clásicos de series de tiempo (AR y MA), posteriormente exploraremos los modelos basados en redes neuronales (LSTM y TCN), y finalmente, analizaremos los modelos fundamentados en árboles de decisión (XGBOOST).

3.1. Mercado FOREX

El mercado de divisas o Foreign Exchange (FOREX), el espacio donde se lleva a cabo la compra y venta de moneda extranjera por parte de entidades de crédito. Estas negociaciones pueden darse de forma directa entre las entidades o con la intermediación de un trader.

El valor de las divisas se establece de manera libre a través de la interacción entre la oferta y la demanda. De esta manera, si hay muchas empresas interesadas en vender una moneda pero pocas desean comprarla, su valor disminuirá. Por el contrario, si existe un alto interés en comprar una moneda y pocas personas están dispuestas a venderla, su valor aumentará.

En este mercado, las transacciones se realizan electrónicamente y de manera bilateral, lo que lo convierte en un mercado sin supervisión. Esta modalidad se conoce en inglés como

OTC (Over the Counter). Esto hace que tenga muchos centros operativos alrededor del mundo, ciudades como Nueva York, Londres y Hong Kong son algunas de las más conocidas y las que mayor volumen de transacciones manejan [6] [3].

Este mercado funciona 24 horas por 5.5 días a la semana, empezando sus labores el domingo en la tarde (19:00 UTC-05), cuando abre el mercado Australiano y acabando el viernes (16:00 UTC-05), cuando cierra el mercado en Nueva York² [6]. Dentro de este mercado hay varios tipos de operaciones que se pueden realizar, desde Spots, que son operaciones ‘de contado’ donde el dinero se transfiere dentro de un plazo máximo de 2 días; Forwards, donde la operación se agenda un día pero se hace efectiva en una fecha futura acordada entre las partes, hasta Swaps, donde se acuerdan las condiciones de compra y recompra de divisas dentro de un plazo determinado, entre otras más.[10].

3.2. Modelos Estadísticos

3.2.1. Modelos de series de tiempo

Los modelos de series de tiempo han sido ampliamente utilizados en el ámbito económico durante varios años, dentro de los primeros libros al respecto resaltan [11],[12] y [13] donde se abordaban las metodologías univariadas, años más tarde en [14] y [15] se exploró el análisis de series de tiempo multivariadas. Teniendo en cuenta este gran despliegue, para este trabajo se considerarán un modelo *clásico AR o MA y uno ARMAX* como punto de referencia para comparar el rendimiento y la precisión de los modelos propuestos.

El modelo más simple e intuitivo es el proveniente de procesos autorregresivos, estos se basan en una estructura de dependencia temporal donde lo que ocurre en el tiempo t depende linealmente del valor observado en el tiempo inmediatamente anterior $t - 1$. Naturalmente, esta lógica de dependencia puede extenderse a p retardos anteriores, lo que da lugar a los procesos autorregresivos de orden p [12].

Modelo AR

Sea y_t un proceso autorregresivo de orden p ($y_t \sim AR(p)$), con $p \in \mathbb{Z}^+$, si para todo $t \in \mathbb{Z}$

²Tener en cuenta los cambios por horarios de verano

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + Z_t \quad (1)$$

donde $c, \phi_1, \dots, \phi_p \in \mathbb{R}$ y $Z_t \sim RB(0, \sigma^2)$. Estos procesos se denominan de *memoria larga* porque la observación en el tiempo t está correlacionada con los acontecimientos de todos los tiempos anteriores. Sin embargo, no todas las series exhiben este comportamiento, y algunas requieren una *memoria de corto plazo* en la que los eventos pasados pero cercanos tengan un impacto mayor que los ocurridos hace mucho tiempo. Estos procesos se conocen como procesos de promedio móvil [13].

Modelo MA

Sea y_t un proceso de promedio móvil (MA) de orden q , con $q \in \mathbb{Z}^+$, si para todo $t \in \mathbb{Z}$

$$y_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2)$$

donde $\theta_1, \dots, \theta_q$ son constantes y $Z_t \sim RB(0, \sigma^2)$ [15].

Ahora, bajo la óptima multivariada, y siguiendo a [Brockwell and Davis](#) en [14], sea $\mathbf{Y}_t = (Y_{t1}, Y_{t2}, \dots, Y_{tm})'$ una series de vectores aleatorios para cada $t \in \mathbb{Z}$, todos definidos sobre el mismo espacio de probabilidad. Su función de medias y matriz de covarianzas están definidas así:

$$\boldsymbol{\mu} = E(\mathbf{Y}_t) = \begin{bmatrix} E(Y_{t1}) \\ E(Y_{t2}) \\ \vdots \\ E(Y_{tm}) \end{bmatrix} \quad (3)$$

$$\Gamma(t+h, t) := Cov(\mathbf{Y}_{t+h}, \mathbf{Y}_t) = \begin{bmatrix} Cov(Y_{t+h,1}, Y_{t1}) & \dots & Cov(Y_{t+h,1}, Y_{tm}) \\ Cov(Y_{t+h,2}, Y_{t1}) & \dots & Cov(Y_{t+h,2}, Y_{tm}) \\ \vdots & \ddots & \vdots \\ Cov(Y_{t+h,m}, Y_{t1}) & \dots & Cov(Y_{t+h,m}, Y_{tm}) \end{bmatrix} \quad (4)$$

$\{\mathbf{Y}_t\}$ será un proceso estacionario en sentido débil si sus momentos $\boldsymbol{\mu}_t$ y $\Gamma(t+h, t)$ son inde-

pendientes de $t, \forall t \in \mathbb{Z}$ y por tanto, $\mu_t = \mu$ y $\Gamma(t+h, t) = \gamma(h)$

Adicionalmente, [14] también define la función de autocorrelación del proceso $\{\mathbf{Y}_t\}$ como $\rho_{ij}(h) = \gamma_{ij}(h)(\gamma_{ii}(0)\gamma_{jj}(0))^{-1/2}$, donde $\rho_{ij}(h) = \text{Corr}(Y_{i,t+h}, Y_{j,t})$.

Teniendo en cuenta los conocimientos previos, se considera ahora un proceso vectorial autorregresivo o VAR por sus siglas en inglés. [15] expresa este proceso como una combinación lineal de los valores pasados de la variable más el componente de error.

Modelo VAR

Sea $\{\mathbf{Y}_t\}$ un proceso k -dimensional tal que $\forall t \in \mathbb{Z}$ se tiene:

$$\mathbf{Y}_t = \mathbf{v} + A_1 \mathbf{Y}_{t-1} + \dots + A_p \mathbf{Y}_{t-p} + \mathbf{U}_t \quad (5)$$

donde $\mathbf{U}_t \sim RB(\mathbf{0}, \Sigma_u)$, Σ_u es no singular, \mathbb{R}^k , A_1, \dots, A_p son matrices reales de tamaño $K \times K$ y $p \in \mathbb{N}$.

Para obtener un proceso autorregresivo (AR) es necesario asumir que los errores U_t son no correlacionados para diferentes valores de t , es decir, $U_t \neq U_s$ para $t \neq s$. Así, es posible afirmar que a la hora de hacer un pronóstico, no se tendrán errores sistemáticos.

Cualquier proceso VAR(p) puede ser expresado como un VAR(1), esto matricialmente se ve como $\mathbf{Y}_t = (Y'_t, Y'_{t-1}, \dots, Y'_{t-(p-1)})_{kpx1}, \mathbf{v} = (v', 0', \dots, 0')_{kpx1}, \mathbf{U} = (u', 0', \dots, 0')_{kpx1}$

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{bmatrix}_{kpxkp}$$

Así pues, este proceso VAR(p) será *estable* solo si todos los valores propios de \mathbf{A} tienen módulo menor a 1, es decir, si el determinante de $(I_{Kp} - \mathbf{A}z)$ es distinto de 0 para todo $|z| \leq 1$.

Adicionalmente, es importante resaltar que un proceso VAR(p) estable es estacionario pero un proceso no estable no es necesariamente no estacionario.

3.3. Modelos de Aprendizaje Automático

Dentro del ámbito del aprendizaje automático, resalta un método conocido como redes neuronales. Su prominencia se debe tanto a su amplio rango de aplicaciones como a su antigüedad. A pesar de que su nombre puede sonar moderno, este método comenzó a gestarse hace varias décadas.

En 1943, Warren S. McCulloch y Walter Pitts publicaron "A Logical Calculus of the Ideas Immanent in Nervous Activity"[16], una teoría que intentaba explicar cómo el cerebro humano podía capturar y comprender patrones complejos a través de conexiones neuronales. Quince años más tarde, Frank Rosenblatt publicó "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain"[17], donde presentó el desarrollo del perceptrón, la forma más básica de red neuronal. Este perceptrón podía reconocer patrones durante su fase de entrenamiento y luego clasificar correctamente entre 2 categorías. Sin embargo, tenía problemas de generalización y no era eficiente cuando la separación entre categorías no era lineal.

Con el paso de los años, se desarrollaron modelos nuevos y mejorados que podían aprender de funciones no lineales e incluso utilizar el método de *backpropagation*, conocido a mediados de los años 80. Todos estos avances sentaron las bases de los diversos tipos de redes neuronales conocidos en la actualidad. Algunos ejemplos incluyen las redes feedforward o perceptrones multicapa (MLP por sus siglas en inglés), donde la información solo fluye hacia adelante. Estas redes constan de una capa de entrada, una de salida y una o más capas intermedias ocultas, cuyas neuronas utilizan la función de activación sigmoide. También encontramos las redes neuronales convolucionales (CNN), ampliamente utilizadas en análisis de imágenes y reconocimiento de patrones, y las redes neuronales recurrentes (RNN). A diferencia de las anteriores, las RNN poseen bucles que permiten la retroalimentación y, por lo tanto, la retención de información. Debido a esta característica, son ampliamente utilizadas en el análisis de series temporales, entre otros usos [18].

3.3.1. Memoria de Largo y Corto Plazo (LSTM)

Dentro del enfoque de este trabajo, centrado en el análisis y pronóstico de datos secuenciales, se da preferencia al uso de redes neuronales recurrentes, en particular a una de sus subclases: las redes LSTM o de memoria de largo y corto plazo.

El desarrollo de las LSTM se inició en 1997 con Hochreiter y Schmidhuber, quienes introdujeron las compuertas el "vector de memoria" [1, 19]. La motivación central de esta estructura radica en permitir que la red aprenda a *olvidar* lo que considere poco relevante y a *recordar* lo que proporcione información más sustancial, incluso si ocurrió en un momento anterior.

Olah detalla la estructura básica de este tipo de red [1], como se muestra en la figura 1. El rectángulo verde representa una capa de la red neuronal, el círculo rosado denota una operación específica y las flechas ilustran el flujo de vectores a través de los componentes de la capa.

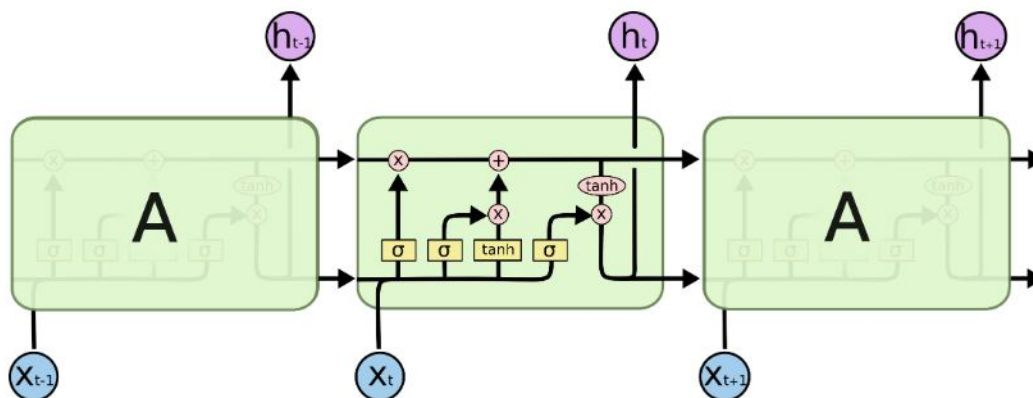


Figura 1: Estructura básica de una red LSTM

Fuente: [1]

En [19], se detallan minuciosamente los componentes de la celda LSTM. Dados x_t y h_{t-1} como las entradas de la celda (la primera es la observación en el tiempo t y la segunda es la salida de la capa LSTM en el tiempo $t - 1$), ambas atraviesan la *compuerta de olvido*. Esta compuerta, implementada mediante una función de activación sigmoideal, genera un valor entre 0 y 1, que determina la cantidad de información a recordar, donde 1 significa recordar todo y 0 olvidar todo. Este valor se multiplica por el vector de estado C_{t-1} , que almacena la memoria de la red hasta el tiempo $t - 1$, para determinar qué información se debe olvidar.

Simultáneamente, la *compuerta de recuerdo* se activa. Está compuesta por dos partes: una utiliza una función de activación sigmooidal para indicar cuánto se debe recordar, y la otra aplica una función tangente hiperbólica (*tanh*) para generar un candidato al vector de estado \tilde{C}_t . Al combinar los resultados anteriores, [1] señala que se establece la información nueva que se añadirá al vector de estado y la que se actualizará. Esto da como resultado un nuevo vector C_t , que antes de pasar a la siguiente capa, debe pasar por un proceso de filtrado mediante una función sigmooidal y luego una transformación con *tanh* para que sus valores queden en el rango entre -1 y 1.

Matemáticamente, se expresa de la siguiente forma:

$$\begin{aligned}
 f_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C * [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W_o * [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned}
 \tag{6}$$

Y gráficamente, tal como se observa en la figura 2

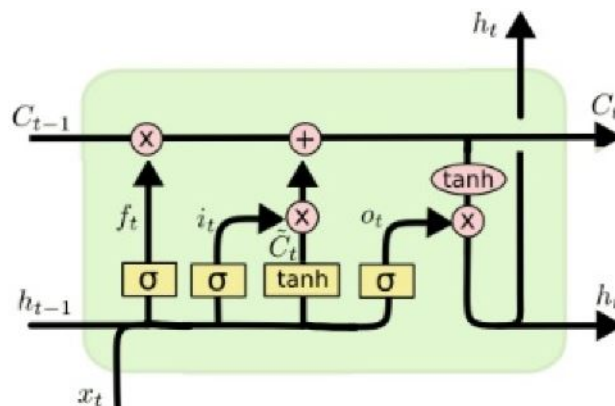


Figura 2: Detalle de una capa de LSTM

Fuente: [1]

3.3.2. Temporal Convolutional Networks (TCN)

Las Redes Convolucionales Temporales o Temporal Convolutional Networks (TCN), como su nombre lo dice, son un tipo de arquitectura utilizadas para modelar secuencias temporales. A diferencia de las redes recurrentes, que utilizan conexiones recurrentes para capturar la información temporal, las TCN emplean capas convolucionales para procesar la secuencia de entrada. La estructura de una TCN consiste en una serie de capas convolucionales dilatadas, donde la dilatación controla la cantidad de pasos de tiempo que cada filtro convolucional considera en su campo receptivo.

Este tipo de arquitectura fue presentada por [Bai et al.](#), donde enfatizan en las dos diferencias claves de una TCN con las redes convolucionales tradicionales: en primer lugar, las convoluciones en su arquitectura son causales, lo que implica que no hay transferencia de información del futuro al pasado, y en segundo lugar, la arquitectura tiene la capacidad de procesar secuencias de longitud variable y generar una secuencia de salida de la misma longitud, similar a las RNN. Adicionalmente, se resalta la habilidad que tienen las redes de examinar extensivamente en el pasado para realizar predicciones, esto gracias a la combinación de redes profundas, que incluyen capas residuales y convoluciones dilatadas.

Para lograr las características mencionadas, las redes están compuestas por capas convolucionales causales. Estas capas, también conocidas como capas completamente conectadas, garantizan que un elemento en la secuencia de salida dependa únicamente de los elementos que lo preceden en la secuencia de entrada. Además, para mantener la misma longitud entre la secuencia de salida y la de entrada, se emplea un relleno de ceros. No obstante, es fundamental aplicar el relleno de ceros exclusivamente en el lado izquierdo del tensor de entrada para asegurar una convolución causal. De esta manera, se evita cualquier influencia de la información futura en la generación de la secuencia de salida.

Con el fin de obtener un información de un pasado distante, es decir que el modelo logre explorar un historial amplio de las series objetivo, se implementa el concepto de dilatación dentro de las capas convolucionales, entendiendo dilatación como la distancia entre los elementos de la secuencia de entrada que se utilizan para calcular una entrada de la secuencia

de salida. Matemáticamente, la operación de dilatación F fue expresada por [20] así:

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i)x_{s-di} \quad (7)$$

siendo x una secuencia de entrada, s un elemento de la secuencia, $f : 0, \dots, k-1 \mapsto \mathbb{R}$ un filtro, d el factor de dilatación y k el tamaño del kernel. Tomando como referencia la figura 3 se comprende mejor el efecto que tiene la dilatación en la obtención de información histórica, toda vez que al aumentar d , aumenta la cantidad de pasos hacia atrás que toma el modelo para generar la siguiente capa convolucional.

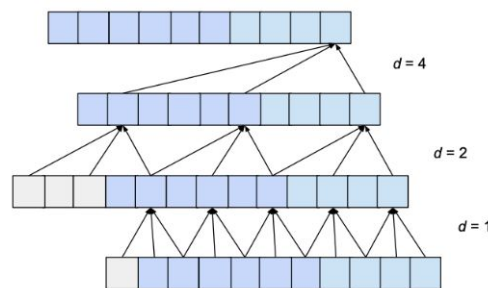


Figura 3: Cómo se ve la dilatación

Fuente: [2]

En [2], se considera además la creación de Bloques residuales, que se entienden como 2 capas convolucionales con el mismo factor de dilatación unidas a la función de activación, normalización y regularización. Estos bloques estarán determinados por el tamaño del kernel k y el factor de dilatación d . En la figura 4 provista por el mismo autor, se evidencia de forma más intuitiva la estructura de este tipo de redes.

3.3.3. Extreme Gradient Boosting (XGBOOST)

Dentro del campo del aprendizaje automático, además de los métodos basados en redes neuronales, se encuentran enfoques que se apoyan en árboles de decisión. Estos enfoques se destacan por su estructura intuitiva, su capacidad de comprensión y su facilidad de visualización, lo que los convierte en opciones ampliamente utilizadas y de fácil comprensión. A diferencia de los modelos de redes neuronales, cuya complejidad interna los convierte en "cajas negras" con un funcionamiento interno poco transparente, los modelos basados en árboles de decisión ofrecen mayor claridad y facilidad de interpretación.

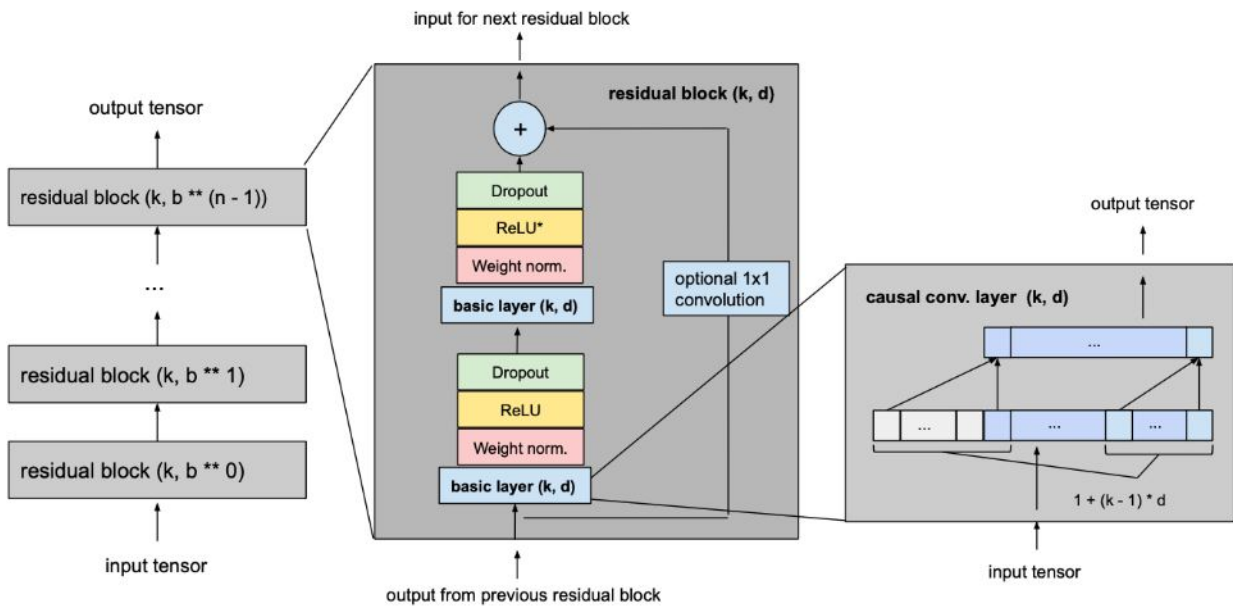


Figura 4: Estructura general de una TCN

Fuente: [2]

Dentro de los más conocidos se encuentra el método de XGBOOST o extreme gradient boosting, introducido por [Chen and Guestrin](#), donde se combinan varios árboles de decisión débiles para formar un modelo robusto. Un árbol de decisión débil es aquel que tiene una capacidad de predicción apenas mejor que la predicción aleatoria. Sin embargo, cuando se unen varios modelos débiles, se puede formar un modelo robusto que sea mucho más preciso. Cada árbol se entrena con una submuestra de los datos, y las predicciones de cada árbol se combinan para formar la predicción final. XGBOOST utiliza un enfoque iterativo, que implica la interacción continua de los árboles dentro del modelo, ajustando los residuos del modelo anterior para refinar la precisión del modelo en cada iteración, capturando patrones complejos en los datos.

En [21] se establece que el objetivo del algoritmo es encontrar los pesos óptimos para cada árbol que minimicen una función de pérdida regularizada. Esta función de pérdida está compuesta por dos términos: el término de pérdida de datos, que mide la discrepancia entre las salidas de predicción y las salidas reales en el conjunto de datos de entrenamiento, y el término de penalización, que mide la complejidad del modelo de ensamblaje de árboles. Este último se controla mediante un parámetro de regularización que penaliza los árboles de regresión con muchas hojas.

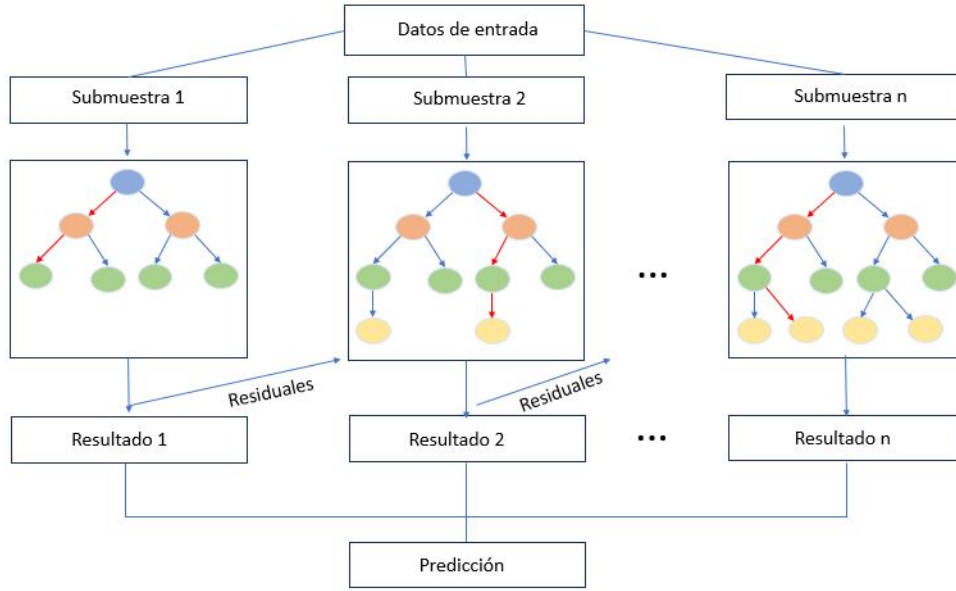


Figura 5: Estructura XGBOOST

Elaboración propia

Formalmente, la función de pérdida se puede expresar de la forma

$$\mathcal{L}^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (8)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

donde \hat{y}_i^t es la predicción en la i -ésima instancia del t -ésimo árbol, l es una función de pérdida convexa y diferenciable que mide la diferencia entre la predicción y el valor real (y_i), y Ω es la penalización se aplica a la complejidad del modelo.

Dada la inmensa cantidad de opciones que hay para la construcción de los árboles, [21] presenta una función de score que permite medir la calidad de la estructura de cada árbol, calculando el valor óptimo de \mathcal{L} , aquí entre menor sea el valor, mucho mejor.

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (10)$$

donde q es la estructura de árbol, I_j el conjunto de instancias de la hoja j , T el número total de hojas en cada árbol y g_i, h_i la primera y segunda derivada de la función de pérdida.

3.4. Métricas de evaluación

Con el fin de comparar los resultados de los distintos modelos, se hará uso de métricas tradicionales para el cálculo de errores de pronóstico, como la raíz del error cuadrático medio (RMSE), error absoluto medio (MAE) y el error porcentual absoluto medio (MAPE). La minimización de estas métricas es fundamental, ya que implica una mayor precisión en las predicciones realizadas por los modelos al reducir el error en los pronósticos.

Adicionalmente, se aplicará el test de Diebold-Mariano ³, que permite comparar la precisión de los dos modelos y determinar si ésta es estadísticamente diferente [22]. Siendo y_i la observación i -ésima, \hat{y}_i el pronóstico i -ésimo y n el tamaño de muestra, dichas estadísticas se calculan como sigue

$$\begin{aligned} \blacksquare \text{ RMSE} &: \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} & \blacksquare \text{ MAPE} &: \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) * 100 \\ \blacksquare \text{ MAE} &: \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \end{aligned}$$

4. Estado del arte

Durante los últimos años, se ha evidenciado un aumento en la aplicación de técnicas de aprendizaje de máquina e inteligencia artificial en el cálculo de pronósticos de series financieras. En 2014, [Khadjeh Nassirtoussi et al.](#) demostró que mediante un algoritmo multicapa aplicado a minería de texto podía utilizar los titulares de noticias como insumo para predecir los movimientos intra-día que presenta la tasa de cambio Euro/Dólar con una certeza del 83.33%. Por otro lado, [19], expone mediante comparación de modelos estadísticos como los autoregresivos, de ruido blanco, SARIMA y los modelos switching Markov versus las redes LSTM, que éstas últimas brindan un mejor pronóstico de la inflación estadounidense, al tener una arquitectura más flexible que les permite capturar la no-linealidad de la serie objetivo.

³Siendo e_{1t} y e_{2t} los errores de pronóstico de dos modelos y $g(\cdot)$ una función de pérdida, sea $d_t = g(e_{1t}) - g(e_{2t})$. Con esto se busca probar la hipótesis nula $H_0 : E(d_t) = 0 \forall t$ vs la hipótesis alterna $H_1 : E(d_t) \neq 0$. Para mayor detalle, ver [22].

Otro de los varios algoritmos utilizados en este campo, es el *variational mode decomposition* (VMD), una herramienta útil en el procesamiento de señales que fue aplicada por [23] como base para el pronóstico de las tasa de cambio de Dólar/Yuan Chino y Dólar/Yen Japonés, obteniendo mejores resultados cuando es anidada a una red LSTM, que cuando se usa directamente el VMD.

Así pues, aunque se ha llevado a cabo muchos estudios respecto a variables macroeconómicas y financieras, sobre todo en relacionadas con la tasa de cambio y con el mercado accionario, los hallazgos han sido diversos. Por ejemplo en [24] y en [25], se buscaba pronosticar el valor de mercado de acciones mediante el uso de redes neuronales, en el primero, la precisión del pronóstico superó los parámetros de referencia pero en el segundo, los resultados no fueron los esperados y a pesar de tener un porcentaje de certeza alto, no logró superar las expectativas.

Algo similar se puede observar en los documentos de [26] y [27], donde el objeto de estudio también es el pronosticar precio de diferentes acciones, pero con enfoques distintos, en el primero el objetivo era aplicar un modelo TCN con módulo de atención y compararlo con otros modelos para determinar cuál se comportaba mejor a la hora de pronosticar, mientras que el segundo, buscaba pronosticar el cambio en el precio de acciones usando datos de alta frecuencia. Como resultado, se determinó que en ambos casos los modelos TCN brindaban mejores pronósticos respecto a otras arquitecturas, aun cuando se hace hincapié en la necesidad de continuar trabajando con conjunto de datos más complejos para poder generalizar las conclusiones.

Lo anterior demuestra el amplio espectro que hay en el ámbito de pronóstico de series financieras, tanto en técnicas como en variables objetivo. Sin embargo, resalta el hecho de que respecto al análisis de volumetría de las transacciones del mercado FOREX no hay mucho desarrollo documentado, a pesar de ser una herramienta útil para los traders y para los clientes a la hora de tomar decisiones de inversión, y en eso está enfocado este trabajo.

5. Metodología

5.1. Recolección de datos

Para el desarrollo del presente trabajo, la variable a pronosticar será extraída de los registros diarios de transacciones del mercado interbancario de divisas, que presenta el Banco de la República en su página web oficial, estos a su vez obtenidos de SET-FX⁴ y de la Superintendencia financiera de Colombia.⁵

Adicionalmente, se contará con las variables auxiliares citadas en el cuadro 1. Dada la facilidad de obtener datos históricos, el análisis se realizará con 2015 a 2022 como periodo de referencia, así se podrá modelar el efecto de la crisis de 2018 y la pandemia de COVID-19 que comenzó en 2020.

Se aclara también, que se tendrá como base el calendario del mercado interbancario, es decir, solo se contarán días donde haya habido operaciones interbancarias, excluyendo fines de semana o festivos internacionales.

Variable	Fuente	Sigla
Volumen transacciones mercado forex	Banco de la República	Vol
Monto transacciones mercado forex	Banco de la República	Monto
Precio del petróleo BRENT (USD)	Investing.com	Brent
Tasa de cambio COP/USD	Banco de la República	TRM
Tasa interés de Política Monetaria del BR	Banco de la República	TPM
Tasa interés de Política Monetaria de la FED	Reserva Federal EE.UU	FRB
Índice DXY	Investing.com	DXY
IPC Total	Banco de la República	IPC
Remesas	Banco de la República	Remesas
Exportaciones	DANE	Export
Importaciones	DANE	Import
Producto Interno Bruto de Colombia	DANE	PIB
Producto Interno Bruto de Estados Unidos	OCDE	GDP
Rendimiento Bono Colombia 10 años	Investing.com	Bono10Y
Índice de volatilidad CBOE VIX	Investing.com	VIX

Cuadro 1: Lista de variables

⁴Sistema electrónico a través del cual los intermediarios del mercado cambiario negocian y registran sus operaciones. Para más detalle <https://set-icap.com/set-icap-fx/acerca-de/>

⁵Para mayor información visite: <https://www.banrep.gov.co/es/estadisticas/mercado-interbancario-divisas-reporte-diario>

5.2. Preprocesamiento de datos

En esta sección, se describen las etapas de procesamiento de los datos utilizados en este estudio, empezando por alinear las series a una misma frecuencia (transformando datos trimestrales o mensuales a diarios), pasando por el manejo de fechas especiales, su respectiva estandarización y finalizando con el ordenamiento de la serie para ser ingresada a cada modelo.

Las variables objetivo en este estudio se refieren a datos diarios relacionados que provienen del mercado FOREX. Por lo tanto, es esencial considerar que el funcionamiento de dicho mercado está sujeto a restricciones específicas. Un ejemplo de ello es su operación exclusiva en días hábiles, excluyendo los días festivos tanto en Colombia como en Estados Unidos. Durante estos días, ya sea en uno de estos países o en ambos, se observa una reducción significativa en el volumen de transacciones. Esta disminución en el volumen genera datos que pueden ser considerados atípicos dentro de las series de tiempo que deben ser identificados y eliminados durante el proceso de limpieza de datos. De esta manera, se garantiza la integridad de los datos utilizados en el análisis y en los modelos de pronóstico.

Después de realizar la limpieza de datos para abordar los efectos de los días festivos, se transformaron en algunas covariables ya que se presentan en una frecuencia distinta a la de la variable dependiente. En el caso de las variables FRB, TPM e IPC, que mantienen un valor constante durante un periodo específico, se asignó dicho valor a cada día del mes correspondiente. Por ejemplo, si el IPC mensual de marzo de 2015 fue de 4,56, este valor se aplicó a cada día de dicho mes. Por otro lado, existen otras variables, como Remesas, Importados, Exportados, PIB de Estados Unidos (GDP) y de Colombia (PIB), cuyos datos corresponden a la acumulación mensual o trimestral. Para estas variables, se utilizó el promedio mensual o trimestral para asignarlos a los días correspondientes. De esta manera, si el valor mensual de las remesas que ingresaron al país en enero de 2015 fue de \$340 millones de dólares, y el mercado estuvo activo durante 19 días en ese mes, el valor diario de ingreso de remesas sería de \$17,89 millones de dólares por día.⁶

⁶Se realizó una revisión de los modelos de frecuencia mixta (MIDAS), sin embargo, en este caso las frecuencias se encuentran invertidas, lo que implica que su implementación requeriría una evaluación más exhaustiva y un tiempo de programación que excede los límites de este trabajo. Por lo tanto, para aplicar el modelo MIDAS,

Como se evidencia en las gráficas 6 y 7, las series objetivo resultantes son muy volátiles lo que dificulta la identificación de tendencias o patrones significativos en ellas. Así, siguiendo el procedimiento visto en [28] y las recomendaciones de [9] sobre el uso frecuente de medias móviles en el análisis técnico, se realizan dos suavizamientos utilizando promedios móviles.

El primero emplea una ventana de 5 días para lograr un suavizamiento a corto plazo equivalente a una semana, mientras que el segundo utiliza una ventana de 40 días para lograr un suavizamiento a largo plazo equivalente a dos meses. Esta estrategia permite observar el comportamiento suavizado de las series en diferentes horizontes temporales. El promedio móvil más sencillo, le asigna un peso de $\frac{1}{N}$ a las N observaciones más recientes y 0 al resto de observaciones, de forma que el promedio móvil M_T se puede expresar como muestra la ecuación 11. Los resultados de este suavizamiento también se pueden ver en la figuras 6 y 7.

$$M_T = \frac{1}{N} \sum_{t=T-N+1}^T y_t \quad (11)$$

Después de aplicar el suavizado a las series objetivo, se lleva a cabo la estandarización de todas las variables, tanto las objetivo como a las covariables. Esta etapa de estandarización fue necesaria debido a la disparidad de escalas presentes en las variables, algunas denotaban cantidad, otras estaban en millones de dólares y otras tantas eran índices. Dicha estandarización fue fundamental para garantizar una adecuada comparabilidad entre las variables y permitir su inclusión en los modelos de análisis de series de tiempo para pronóstico.

sería necesario desarrollarlo desde cero, debido a que la frecuencia de la variable dependiente es más alta que la de las variables independientes y el enfoque convencional del modelo MIDAS está diseñado al contrario.

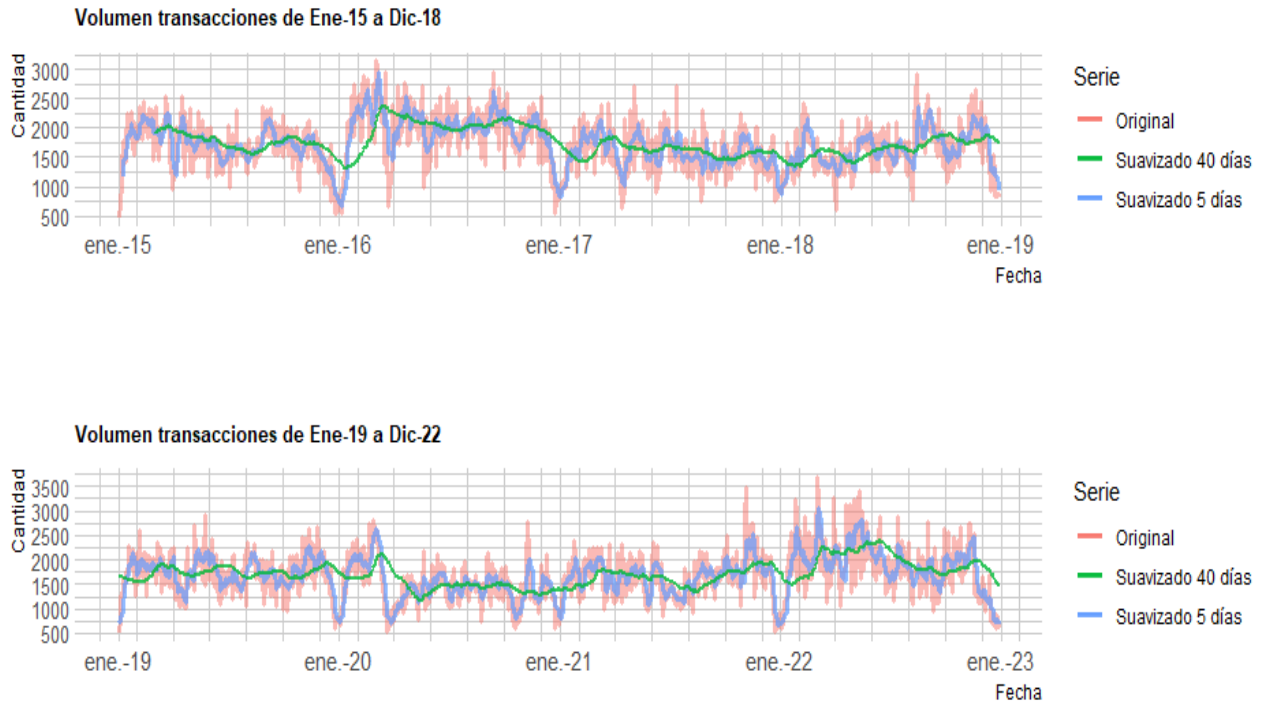


Figura 6: Volumen transacciones - Suavizado

5.3. Validación cruzada

La validación cruzada, también conocida como cross-validation, es una técnica ampliamente utilizada en el campo del aprendizaje automático para evaluar y validar los resultados del modelo, evitando que los resultados sean influenciados por la división específica de los datos utilizados. Esta técnica implica la partición del conjunto de datos en dos subconjuntos: uno destinado al entrenamiento del modelo y otro para la validación. De manera iterativa, el modelo se entrena y valida utilizando cada uno de estos subconjuntos, garantizando así su capacidad de generalización. El enfoque comúnmente empleado es el método de validación cruzada k-fold, donde el conjunto de datos se divide en k subconjuntos aleatorios, y se llevan a cabo k iteraciones de validación. Esta técnica permite una evaluación exhaustiva y robusta del rendimiento del modelo al utilizar diferentes combinaciones de datos para entrenamiento y validación.

Sin embargo, para el caso de series temporales, esa partición del conjunto de datos no se pue-



Figura 7: Monto negociado - Suavizado

de hacer de forma aleatoria ya que se estaría omitiendo que existe una relación temporal en los datos, así pues, tal como se expone en [29] la partición se hace tomando como conjunto de entrenamiento las observaciones que ocurrieron antes de aquellas que se consideren como conjunto de prueba, tal como muestra la figura 8.

En este trabajo se usarán 10 folds para realizar la validación cruzada, esto para que los periodos donde hay mayor volatilidad estén tanto en el conjunto de entrenamiento como en el de test y así obtener un modelo final más robusto.

5.4. Implementación de los modelos

De acuerdo al objetivo de este trabajo, para medir el desempeño de las técnicas es necesario modelar y pronosticar el monto y volumen usando técnicas básicas de series de tiempo (ARMA y ARMAX), redes neuronales tipo LSTM y TCN, y métodos basados en árboles de decisión como XGBOOST. Para todos los modelos se aplicará una validación cruzada

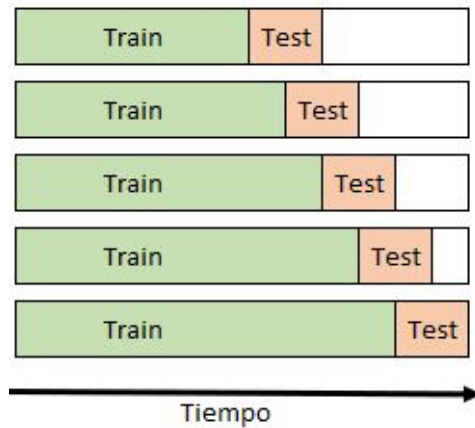


Figura 8: Ejemplo Validación cruzada para series temporales

con 10-folds y se optimizarán los hiperparámetros utilizando *Optuna*⁷. Finalmente, se compararán los resultados obtenidos por cada enfoque para determinar cuál de ellos presenta un mejor rendimiento. El modelamiento y los cálculos se realizarán en los lenguajes de programación R y Python, aprovechando las ventajas que cada uno ofrece para las diferentes tareas.

5.4.1. Modelo Autoregresivo

Para el ajuste de este modelo es necesario primero verificar que las series objetivo muestran el comportamiento de una serie estacionaria; entendiendo estacionariedad como tener media cero y varianza constante. Para comprobar esto se realizan pruebas de raíz unitaria⁸, que permiten establecer si es necesario diferenciar los datos. Para este caso, se aplicaron los test de Dickey-Fuller (ADF), Phillips-Perron (PP) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS), donde para los primeros la hipótesis nula del test es que la serie de tiempo tiene una raíz unitaria, mientras que la hipótesis alternativa es que la serie es estacionaria; para KPSS es al contrario, la hipótesis nula es que la serie de tiempo es estacionaria, mientras que la hipótesis alternativa implica que la serie tiene raíz unitaria. Así, para este caso, si el p-valor asociado a la prueba es menor al nivel de significancia $\alpha = 0,05$, se rechaza la hipótesis nula en todos los casos. En la tabla 2 se encuentran los resultados de los test, de donde se puede concluir que las series suavizadas con una ventana de 5 días no tienen presencia de raíz unitaria por ende se consideran estacionarias, caso contrario sucede con las series suavizadas con ventana

⁷<https://optuna.org/>

⁸Estas pruebas muestran si se presenta una tendencia estocástica en la variable de interés

de 40 días, las cuales deben ser diferenciadas una vez para ser estacionarias.

	ADF Test		PP Test		KPSS Test	
	Estadístico	p-valor	Estadístico	p-valor	Estadístico	p-valor
Volumen 5 días	-7.40	0.01	-118.83	0.01	0.64	0.02
Monto 5 días	-7.90	0.01	-141.77	0.01	0.53	0.03
Volumen 40 días	-3.25	0.08	-18.26	0.10	1.26	0.01
Monto 40 días	-3.08	0.05	-22.47	0.04	1.45	0.01

Cuadro 2: Pruebas raíz unitaria

Se encontró un modelo univariado ARMA(2,2) para la serie de volumen suavizado a 5 días, con criterios de información de Akaike (AIC) y bayesiano (BIC) de -68.99 y -41.56, respectivamente. Asimismo, se identificó un modelo AR(5) para la misma variable suavizada a 40 días (diferenciada), con valores de AIC = -5304.63 y BIC = -5304.58. Los coeficientes relacionados para cada modelo se encuentran en las tablas 5, 6 del anexo 9.1 de modelos autoregresivos univariados.

En relación a las series de monto suavizado con ventana de 5 días, se observa que el modelo ajustado es un ARMA(2,2), con valores de AIC y BIC calculados en 486.09 y 513.53, respectivamente. Por otra parte, para las series de monto suavizado con 40 días (diferenciado), el análisis revela un modelo ARMA(1,4) con criterios de información que indican un AIC de -4586.37 y un BIC de -4553.56. Los coeficientes de ambos modelos se encuentran detallados en la tabla 7 ubicada en el anexo 9.1

Paso seguido, se analiza el comportamiento de las series objetivo cuando se incluyen variables auxiliares al modelo. Inicialmente, se calcula la correlación que existe entre las variables para observar gráficamente cómo se asocian dichas variables. En la Figura 9a, se muestra que la variable *Vol_5d* está fuertemente asociada linealmente con la variable *Monto_5d*, con un coeficiente de correlación de 0.89. La siguiente correlación positiva más alta es de tan solo 0.31 con el IPC. En cuanto a las correlaciones negativas, destacan GDP con -0.04 y TRM con -0.05.

Al analizar las relaciones de *Monto_5d* con el resto de las variables, excluyendo *Vol_5d*, encontramos que el precio del petróleo Brent presenta una correlación positiva de 0.34, mientras

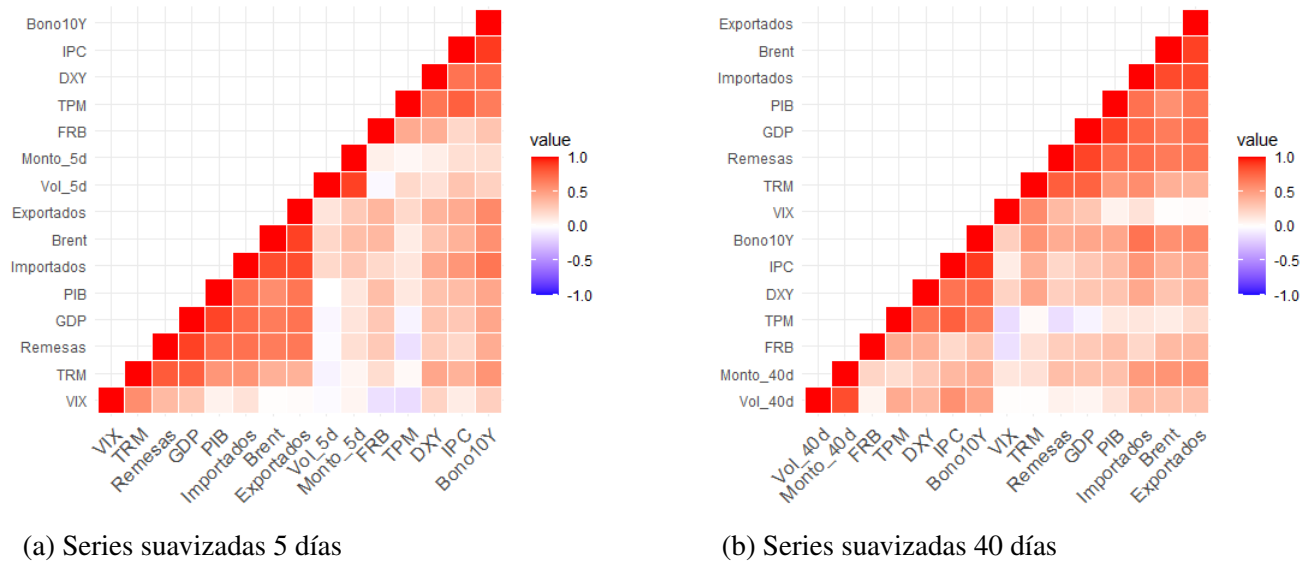


Figura 9: Matrices de correlación

que la tasa de política monetaria del Banco de la República (TPM) tiene un coeficiente de solo 0.04. La TRM y el índice VIX le siguen de cerca, con 0.05 cada uno.

En relación a la Figura 9b, se observa que la variable *Vol_40d* continúa mostrando una fuerte asociación con *Monto_40d*, con un coeficiente de correlación de 0.85, seguida nuevamente por el IPC con 0.57. La relación más débil se presenta con la tasa representativa del mercado y el índice VIX, con un coeficiente de 0.01. En el caso de *Monto_40d*, las variables exógenas con las que se presenta una mayor relación son Exportados con 0.56 y Brent con 0.55, mientras que TRM y VIX son las que presentan una relación menor, con coeficientes de 0.13 y 0.16, respectivamente.

Con la información recopilada, se procede a ajustar los modelos ARMA con variables exógenas para determinar la significancia de aquellas que mostraron una alta correlación. Para ello, se inicia ajustando los modelos con todas las variables exógenas y posteriormente se realiza un refinamiento, eliminando las variables que no resultan relevantes.

Los modelos resultantes para las 4 variables de interés se muestran a detalle en el anexo 9.2 de modelos autoregresivos con variables auxiliares. Cabe resaltar que para todos, ε es un proceso que de ruido blanco con media cero y varianza constante.

5.4.2. Modelo LSTM

Para que los datos puedan ser ingresados correctamente a una red LSTM, es necesario transformarlos en tensores con una forma específica, generalmente de la forma (número de muestras en el lote = batch size, número de pasos de tiempo = sequence length, número de características = features). Esto asegura que los datos se ajusten a la estructura de entrada esperada por la red LSTM y permitan el procesamiento secuencial y la captura de información temporal. En este caso, el valor de batch size y sequence length fue igual para los cuatro modelos, 16 y 20 respectivamente.

Al igual que el modelo autoregresivo, para la validación cruzada se usan 10 folds y se toma como base una red que cuenta con 2 capas⁹ LSTM con función de activación ReLu, una capa de dropout y una capa final completamente conectada. Para cada una de las series objetivo se optimizaron los hiperparámetros de cantidad de nodos en cada capa, valor de dropout y tasa de aprendizaje, adicionalmente se probaron los modelos con distintas covariables para elegir el que menor error de pronóstico arrojará. Así pues, para esta arquitectura, los modelos LSTM quedaron como sigue:

V. objetivo	No. nodos capa 1	No. Nodos capa 2	Dropout	Tasa aprendizaje
Vol 5d	31	38	0.3	0.000048
Monto 5d	61	48	0.1	0.000096
Vol 40d	60	62	0.4	0.000099
Monto 40d	35	15	0.1	0.000093

Cuadro 3: Hiperparámetros LSTM

Se encontraron covariables relevantes que mejoraron el error de pronóstico para cada una de las series objetivo. En el caso de la serie de volumen suavizado a 5 días, se incluyeron las covariables Monto_5d, IPC y Remesas. Para la serie de monto suavizado a 5 días, se utilizaron Vol_5d, TPM, Remesas y Exportados. En cuanto a la serie de volumen suavizado a 40 días, se consideraron Monto_40d, IPC, TPM y GDP. Por último, en el pronóstico de la serie de monto suavizado a 40 días, se incluyeron Vol_40d, Exportados y FRB.

⁹Se decide trabajar con 2 capas en pro de complejizar innecesariamente el modelo.

5.4.3. Modelo TCN

A diferencia del modelo LSTM, para la red TCN no fue necesario transformar la serie a un tensor para que fuese bien recibida por el modelo, solamente se hizo la división por lotes. Si bien se usó la misma partición para la validación cruzada, las series ingresaban al modelo como objetos de series de tiempo.

Respecto a los hiperparámetros, para los modelos cuya variable respuesta era el volumen de transacciones (suavizado tanto 5 como 40 días), se usó un tamaño de lote de 36, un kernel de tamaño 5, 3 filtros, un dropout de 0.1, una dilatación de 2 y una tasa de aprendizaje de 0.001. En el caso de los modelos para la variable Monto (con sus dos suavizamientos), la arquitectura es la misma que para volumen excepto por el tamaño del lote, que pasa a 32.

Cabe resaltar que a pesar de haber probado diferentes combinaciones de variables, para este modelo los mejores resultados se obtuvieron con la inclusión de todas las covariables en los 4 casos.

5.4.4. Modelo XGBOOST

En el caso de XGBOOST, también se utilizó el método de validación cruzada con 10-folds y los datos ingresan como objetos tipo lista. Para los cuatro modelos la arquitectura fue la misma, 1.000 árboles de máximo 3 niveles de profundidad y con una tasa de aprendizaje de 0.01.

A diferencia del modelo anterior, con XGBOOST los modelos con mejores resultados solo involucra algunas de las covariables, de forma que para pronosticar el volumen suavizado 5 días se usan las variables Monto_5d, IPC, Remesas y TPM; para el monto negociado suavizado 5 días se cuenta con Vol_5d, IPC y Remesas; para el volumen suavizado 40 días repiten Monto_40d y IPC; finalmente, para el pronóstico del monto negociado suavizado 40 días se usan las variables Vo_40d, TPM y FRB.

5.5. Optimización con Optuna

Con el propósito de mejorar el rendimiento de los modelos de machine learning, se empleó la biblioteca Optuna¹⁰ para llevar a cabo la optimización de hiperparámetros. A través de Optuna, se buscó identificar la combinación de hiperparámetros que maximizara el desempeño de los modelos minimizando el RMSE.

Para este proceso de optimización, primero se definió un espacio de búsqueda de hiperparámetros 8; luego, se realizaron 100 iteraciones en las cuales se evaluaron y ajustaron los modelos utilizando distintas combinaciones de hiperparámetros. Utilizando algoritmos de búsqueda automáticos, Optuna seleccionó las combinaciones más prometedoras, fundamentándose en los resultados obtenidos durante la validación cruzada.

En cada iteración, Optuna evaluó el rendimiento del modelo empleando el RMSE, y registró los resultados para su posterior análisis. Al concluir el proceso de optimización, se seleccionó la combinación de hiperparámetros que arrojó los mejores resultados y se utilizó para la implementación definitiva de los modelos en la sección 5.4.

6. Resultados

En esta sección, se presentan los resultados obtenidos con la implementación de los modelos propuestos y su evaluación.

Para evaluar el desempeño de los modelos de pronóstico utilizados en este proyecto, se emplearon las tres métricas descritas en la sección 3.4 para medir la precisión de los modelos y comparar su rendimiento. Los resultados obtenidos se presentan en la tabla 4, y que se complementan con las gráficas de pronóstico para 30 días 10 y 11 que evidencian lo consignado en la tabla y proporcionan una representación más intuitiva y facilitan la comprensión de los resultados obtenidos.

En términos de RMSE, el modelo ARMAX obtuvo los mejores resultados en tres de las cua-

¹⁰Mayor detalle sobre Optuna en <https://optuna.org/>

Métrica	RMSE				
Modelos	ARMA	ARMAX	LSTM	TCN	XGBOOST
Vol 5d	1.082	0.504	1.028	1.611	0.403
Monto 5d	0.965	0.336	0.546	1.208	0.429
Vol 40d	0.077	0.046	0.518	0.667	0.384
Monto 40d	0.086	0.053	0.280	0.646	0.424

Métrica	MAE				
Vol 5d	0.823	0.369	0.961	1.150	0.329
Monto 5d	0.751	0.258	0.459	0.978	0.352
Vol 40d	0.062	0.034	0.457	0.569	0.318
Monto 40d	0.068	0.037	0.233	0.548	0.347

Métrica	MAPE				
Vol 5d	96.182	66.290	14.410	257.500	0.817
Monto 5d	101.659	216.824	1.490	111.009	1.869
Vol 40d	102.784	98.838	8.452	569.372	2.430
Monto 40d	98.611	226.164	1.710	1,619.379	2.869

Cuadro 4: Métricas por modelo

tro series evaluadas. Específicamente, para el pronóstico de volumen y monto suavizado 40 días, el modelo ARMAX logró un RMSE de 0.046 y 0.053 respectivamente, superando a los otros modelos. Solamente, en el pronóstico de volumen suavizado 5 días, el modelo XGBOOST mostró un mejor desempeño con un RMSE de 0.403. Esto muestra que a pesar de ser un modelo computacionalmente sencillo, logra capturar mucho mejor las características propias de la serie y se acerca mucho al verdadero proceso generador de los datos.

En cuanto al MAE, el modelo ARMAX también se destacó al obtener los valores más bajos en las métricas de pronóstico de volumen y monto suavizado 40 días y de monto suavizado 5 días, resaltando los dos primero con un MAE de 0.034 y 0.037 respectivamente. Similarmente, el modelo XGBOOST exhibió un rendimiento superior en el pronóstico de volumen suavizado 5 días, alcanzando un MAE de 0.329.

Con relación al MAPE, nuevamente el modelo XGBOOST se destaca al alcanzar el valor más bajo para el volumen suavizado 5 días, podría decirse que es aparentemente el mejor modelo para esta serie objetivo. No obstante, también obtiene el menor valor para el pronóstico de volumen suavizado a 40 días, con un MAPE de 2.430. Por otro lado, para los pronósticos de

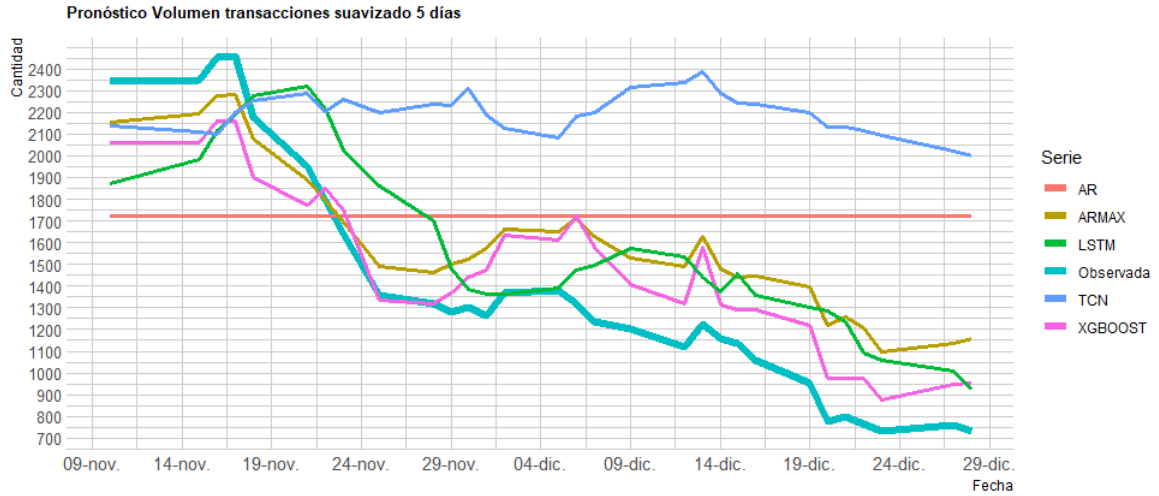
monto suavizado tanto a 5 como a 40 días, el modelo LSTM mostró el mejor desempeño con un MAPE de 1.490 y 1.710 respectivamente.

Si bien los resultados sugieren que los modelos ARMAX y XGBOOST son los mejores para cada una de las series objetivo, es pertinente realizar una prueba estadística para determinar si existen diferencias significativas en cuanto a la precisión de pronóstico. Para este propósito, se emplea el test de Diebold-Mariano¹¹, el cual compara los errores de pronóstico de dos modelos y, a través de una prueba de hipótesis, determina si los pronósticos de los modelos son iguales o cuál de ellos presenta un menor error.

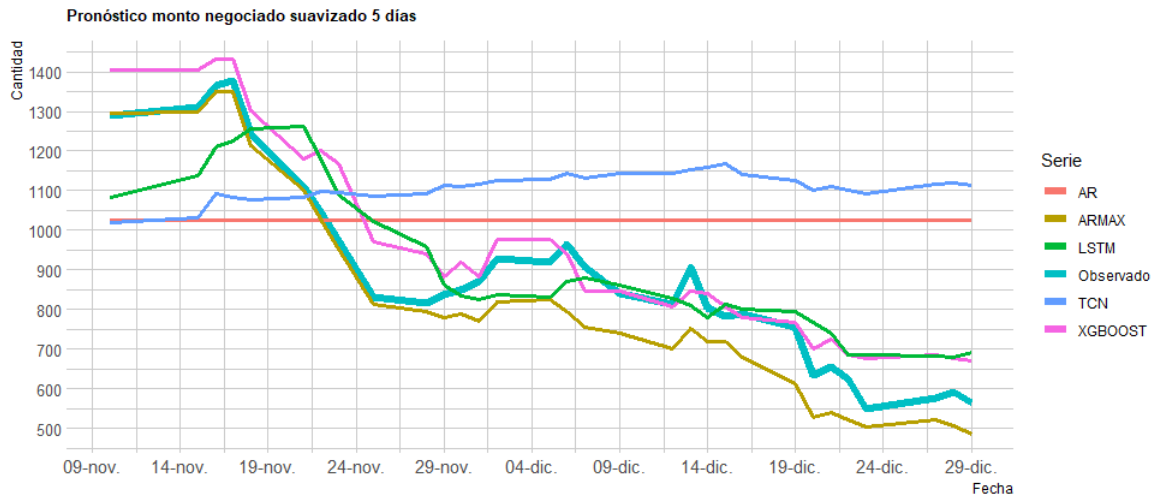
De acuerdo a los resultados del test presentados en las tablas 9, 11, 10 y 12 del anexo 9.4, se puede concluir que el modelo XGBOOST exhibe un rendimiento superior en términos de precisión para el pronóstico de las series suavizadas a corto plazo (5 días). Por su parte, los modelos ARMAX y LSTM también demuestran una precisión estadísticamente similar en esta categoría. En lo que respecta al pronóstico de las series suavizadas a largo plazo (40 días), se destaca el modelo LSTM como el más efectivo, especialmente en el pronóstico del monto negociado, seguido por el modelo XGBOOST como el segundo mejor en esta clasificación.

De igual manera, se ratifica que el modelo ARMA exhibe el peor desempeño en lo que respecta a la precisión del pronóstico. Estos resultados refuerzan la evidencia de que los modelos más avanzados, como ARMAX, LSTM y XGBOOST, superan al modelo ARMA tanto en términos de precisión como de capacidad de pronóstico. Los detalles completos de los resultados de la prueba de Diebold-Mariano están disponibles en el anexo correspondiente (9.4), en las tablas 9, 10, 11, y 12.

¹¹Más detalle de la prueba en el anexo 9.4.

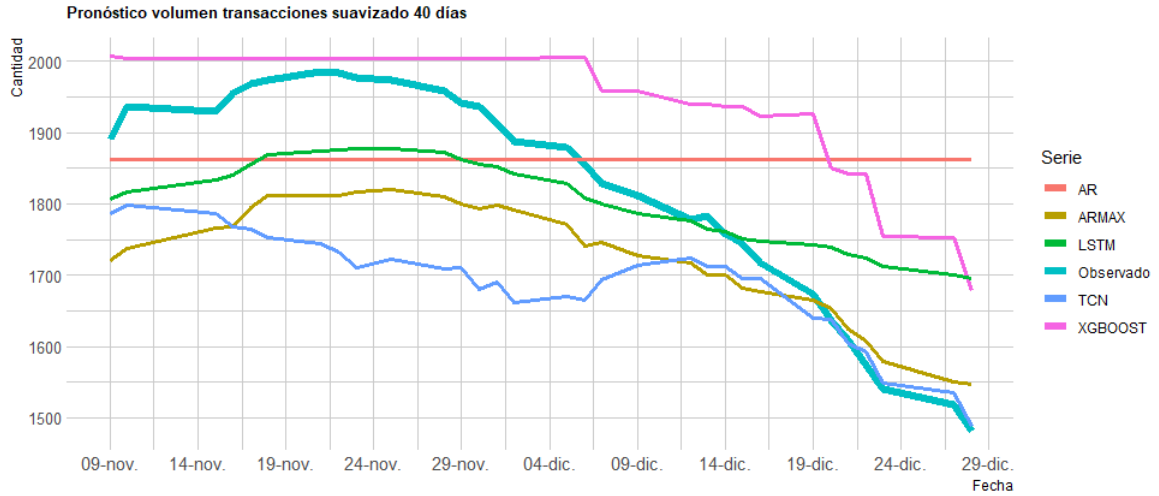


(a) Volumen transacciones

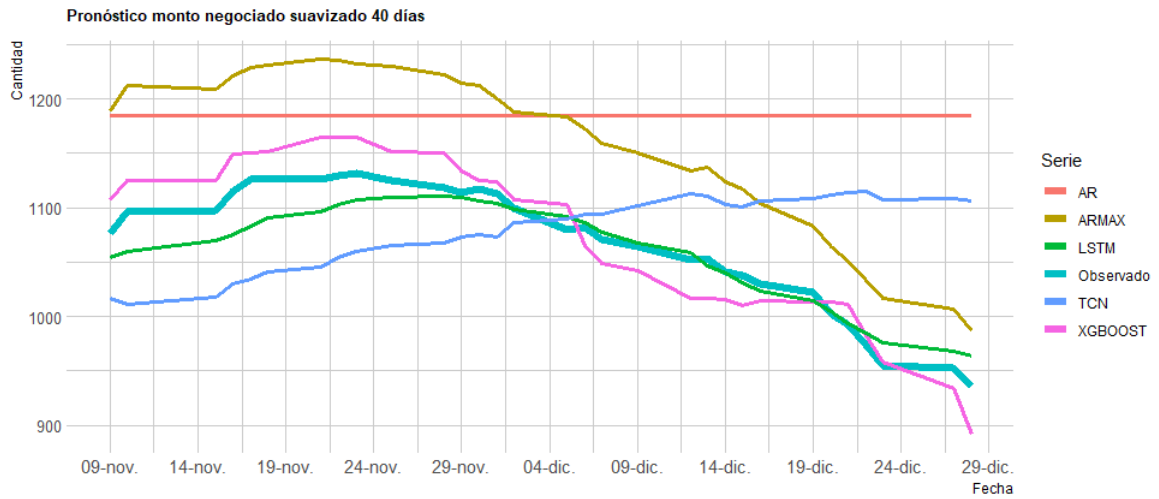


(b) Monto negociado

Figura 10: Pronósticos series suavizadas 5 días



(a) Volumen transacciones



(b) Monto negociado

Figura 11: Pronósticos series suavizadas 40 días

7. Discusión

Siendo XGBOOST el modelo elegido para pronosticar las series suavizadas una semana (5 días), es relevante analizar las covariables que intervienen en cada uno de estos modelos. Para el volumen de transacciones, las covariables relevantes son el IPC, las remesas y el monto negociado, mientras que para pronosticar el Monto se requiere conocer la TPM, el volumen de exportados, las remesas y el volumen de transacciones realizadas.

La inclusión de estas variables cobra sentido al analizar el comportamiento del mercado. Por ejemplo, mayoritariamente, las remesas consisten en transacciones realizadas por personas naturales, que involucran movimientos por montos moderados y mantienen un flujo de divisas constante. Por su parte, el IPC ha mostrado un comportamiento volátil en los últimos años, reflejando los cambios en la economía, lo que lo convierte en una variable de corto plazo que no proporciona muchas pistas sobre el futuro a largo plazo del país.

Además, se encuentra que los exportados aportan información importante para el pronóstico del monto a corto plazo. Esto puede deberse a que cuando los exportadores participan en la jornada, aportan una cantidad significativa de capital, lo que genera la percepción de un mayor volumen de compra/venta de dólares por parte de los agentes del mercado.

Ahora, si nos centramos en las series de largo plazo y sus covariables, con LSTM como modelo elegido, destaca la presencia de la Tasa de Política Monetaria (TPM) del Banco de la República. Esta tasa proporciona información a los inversionistas internacionales sobre el comportamiento a largo plazo de nuestra economía. Cuando la TPM es alta, el país se vuelve atractivo, lo que genera la entrada de divisas. Es común que una estrategia como el *carry trade*¹² cobre relevancia en este escenario, ya que facilita obtener ganancias al invertir en una divisa con alta tasa de interés.

Otra variable relevante que surge del modelado para pronosticar el volumen transaccional suavizado a 40 días, es el GDP. Es esencial considerar las fluctuaciones en el producto interno bruto de Estados Unidos, especialmente para los agentes del mercado FOREX y los

¹²Mayor detalle se puede encontrar en <https://www.banrep.gov.co/es/borrador-957>

traders. Se sabe que si el GDP disminuye durante varios periodos consecutivos, los inversionistas tienden a trasladar sus operaciones a países emergentes que ofrecen perspectivas más favorables. Esto conlleva a una mayor demanda de divisas como el peso colombiano, el real brasileño y el peso mexicano. En contraste, si el GDP aumenta y demuestra constancia, la perspectiva económica en Estados Unidos mejora, lo que resulta en un incremento en la compra de dólares.

Es importante destacar que en todos los casos, el pronóstico del volumen está estrechamente ligado al monto de las transacciones y viceversa, evidenciando la interdependencia entre ambos factores en el mercado FOREX. Estos hallazgos proporcionan información valiosa para los participantes del mercado y permiten comprender mejor los factores económicos y financieros que influyen en las transacciones de divisas.

8. Conclusiones

Este estudio ha permitido determinar el nivel de certeza de los pronósticos de volumen de transacciones USD-COP en el mercado interbancario de divisas. La evaluación de métricas como el RMSE, MAE y MAPE ha proporcionado una comparación precisa de los diferentes modelos y su desempeño relativo.

Entre los modelos evaluados, se encontró que el modelo XGBOOST demostró un excelente desempeño en el pronóstico de volumen a corto plazo (5 días), seguido por los modelos ARMAX y LSTM, cuyos desempeños fueron similares. Para el pronóstico a largo plazo (40 días), el modelo LSTM se destacó, seguido por el modelo XGBOOST.

Los modelos ARMAX, LSTM y XGBOOST demostraron ser más certeros y eficaces en comparación con el modelo ARMA y TCN, superándolos en términos de precisión y capacidad de pronóstico. Aunque el desempeño del ARMA era el esperado, si llamó la atención la capacidad predictiva del modelo TCN ya que no alcanzó las expectativas. Esto resalta la importancia de evaluar y comparar diferentes enfoques para obtener los mejores resultados en cuanto a predicciónse refiere.

Finalmente, estos resultados y los modelos ajustados se convierten en herramientas valiosas para los traders y otros participantes del mercado, brindándoles información confiable para la toma de decisiones. Como trabajo futuro, se recomienda explorar otros modelos y covariables, así como considerar el análisis con un enfoque de frecuencias mixtas para abordar de forma distinta el análisis de mercado.

Referencias

- [1] C. Olah, “Understanding lstm networks,” aug 2015. [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [2] F. Lässig, “Temporal convolutional networks and forecasting,” s.f. [Online]. Available: <https://unit8.com/resources/temporal-convolutional-networks-and-forecasting/>
- [3] B. of International Settlements (BIS), “OTC foreign exchange turnover in April 2022,” Tech. Rep., 2022. [Online]. Available: https://www.bis.org/statistics/rpfx22_fx.htm
- [4] M. Markova, “Foreign exchange rate forecasting by artificial neural networks,” in *APPLICATION OF MATHEMATICS IN TECHNICAL AND NATURAL SCIENCES: 11th International Conference for Promoting the Application of Mathematics in Technical and Natural Sciences - AMiTaNS’19*, vol. 2461, Albena, Bulgaria, 2019. [Online]. Available: <http://aip.scitation.org/doi/abs/10.1063/1.5130812>
- [5] Woon-Seng Gan and Kah-Hwa Ng, “Multivariate FOREX forecasting using artificial neural networks,” in *Proceedings of ICNN’95 - International Conference on Neural Networks*, vol. 2. Perth, WA, Australia: IEEE, 1995, pp. 1018–1022. [Online]. Available: <http://ieeexplore.ieee.org/document/487560/>
- [6] E. Ortega G., “¿los tipos forward pronostican correctamente el tipo de cambio futuro por parte del mercado?” Master’s thesis, ICADE Business School, Julio 2016.
- [7] G. Galati, “Trading volumes, volatility and spreads in FX markets: evidence from emerging market countries,” *BIS Papers*, no. 2, pp. 197–229.
- [8] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, “Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment,” *Expert Systems with Applications*, vol. 42, no. 1, pp. 306–324, Jan. 2015. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0957417414004801>
- [9] J. J. Murphy, *Analisis Tecnico de Los Mercados Financieros (Spanish Edition)*. Gestion 2000, 2000.

- [10] A. C. Corella, “El gran abanico mundial: “Mercado de divisas”,” Master’s thesis, Universidad de Zaragoza, 2016.
- [11] G. E. Box and G. M. Jenkins, *Time series analysis: forecasting and control, revised ed.* Holden-Day, 1976.
- [12] V. M. Guerrero, *Análisis estadístico de series de tiempo económicas.* México,D.F. : International Thomson Editores, S. A, 2003.
- [13] D. Peña, *Análisis de series temporales.* Madrid: Alianza Editorial, 2010.
- [14] P. J. Brockwell and R. A. Davis, *Introduction to time seriesand forecasting, second edition.* New York : Springer, 2002.
- [15] H. Lütkepohl, *New introduction to multiple time series analysis.* Berlin: New York : Springer, 2005.
- [16] P. W. McCulloch, W.S., “A logical calculus of the ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics* 5, p. 115–133, 1943.
- [17] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.” *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.
- [18] I. C. Education, “Neural networks,” 2020. [Online]. Available: <https://www.ibm.com/cloud/learn/neural-networks>
- [19] A. Almosova and N. Andresen, “Nonlinear inflation forecasting with recurrent neural networks,” *Journal of Forecasting*, p. for.2901, Sep. 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/for.2901>
- [20] S. Bai, J. Z. Kolter, and V. Koltun, “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling,” Apr. 2018, arXiv:1803.01271 [cs]. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [21] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, arXiv:1603.02754 [cs]. [Online]. Available: <http://arxiv.org/abs/1603.02754>

- [22] F. X. Diebold and R. S. Mariano, “Comparing Predictive Accuracy,” *Journal of business Economic Statistics*, vol. 13, no. 3, p. 41, 1995. [Online]. Available: <https://www.sas.upenn.edu/~fdiebold/papers/paper68/pa.dm.pdf>
- [23] Y. Huang, Y. Gao, Y. Gan, and M. Ye, “A new financial data forecasting model using genetic algorithm and long short-term memory network,” *Neurocomputing*, vol. 425, pp. 207–218, Feb. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0925231220306718>
- [24] X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, “Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets,” *Mathematical Problems in Engineering*, vol. 2018, pp. 1–11, 2018. [Online]. Available: <https://www.hindawi.com/journals/mpe/2018/4907423/>
- [25] K. S. Gan, K. O. Chin, P. Anthony, and S. V. Chang, “Homogeneous Ensemble FeedForward Neural Network in CIMB Stock Price Forecasting,” in *2018 IEEE International Conference on Artificial Intelligence in Engineering and Technology (ICAIET)*. Kota Kinabalu, Malaysia: IEEE, Nov. 2018, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/8638452/>
- [26] P. Janssen, “Attention based Temporal Convolutional Network for stock price prediction,” Master’s thesis, Utrecht University, April 2022. [Online]. Available: <https://studenttheses.uu.nl/handle/20.500.12932/41588>
- [27] W. Dai, Y. An, and W. Long, “Price change prediction of Ultra high frequency financial data based on temporal convolutional network,” *Procedia Computer Science*, vol. 199, pp. 1177–1183, 2022. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1877050922001508>
- [28] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*, ser. Wiley series in probability and statistics. Hoboken, N.J: Wiley-Interscience, 2008.
- [29] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.

- [30] L. S. . N. P. Harvey, D., “Testing the equality of prediction mean squared errors,” *International Journal of Forecasting*, vol. 13(2), p. 281–291, 1997.

9. Anexos

9.1. Modelos autoregresivos univariados

A continuación se encuentran las salidas de R que muestran el ajuste de los modelos ARMA univariados para cada una de las variables objetivo, así como las tablas a las que se hace referencia en el documento.

	ar1	ar2	ma1	ma2
coef	0.1063	0.7447	1.4127	0.6927
desv. est.	0.0197	0.0213	0.0254	0.0345

Cuadro 5: Coeficientes ARMA(2,2) - volumen suavizado 5 días

```
Series: v5_train
ARIMA(2,0,2) with zero mean

Coefficients:
      ar1      ar2      ma1      ma2
  0.1063  0.7447  1.4127  0.6927
s.e.  0.0197  0.0213  0.0254  0.0345

sigma^2 = 0.05601:  log likelihood = 39.5
AIC=-68.99  AICC=-68.96  BIC=-41.56

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.0007839182 0.2364051 0.1856036 175.4493 267.6127 0.888173 -0.1244686
> coefest(arv5)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 0.106254   0.019665   5.4031 6.551e-08 ***
ar2 0.744703   0.021263  35.0229 < 2.2e-16 ***
ma1 1.412701   0.025400  55.6176 < 2.2e-16 ***
ma2 0.692668   0.034528  20.0610 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 12: Ajuste modelo ARMA volumen suavizado 5 días

	ar1	ar2	ar3	ar4	ar5
coef	0.3773	0.0939	0.0644	0.1438	0.0927
desv. est.	0.0238	0.0252	0.0253	0.0253	0.0238

Cuadro 6: Coeficientes AR(5) - Δ volumen suavizado 40 días

```

Series: vol40_train
ARIMA(5,0,0) with zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5
      0.3774  0.0938  0.0646  0.1436  0.0927
s.e.  0.0238  0.0252  0.0253  0.0253  0.0238

sigma^2 = 0.002824; log likelihood = 2656.33
AIC=-5300.66  AICc=-5300.62  BIC=-5267.86

Training set error measures:
              ME          RMSE          MAE  MPE  MAPE          MASE          ACF1
Training set -8.423672e-05  0.05306852  0.0412892  NaN  Inf  0.8419942  0.001267485
> coeftest(arv40)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.377399   0.023817  15.8456 < 2.2e-16 ***
ar2  0.093842   0.025241   3.7179 0.0002009 ***
ar3  0.064609   0.025308   2.5529 0.0106817 *
ar4  0.143623   0.025266   5.6844 1.313e-08 ***
ar5  0.092715   0.023849   3.8876 0.0001012 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 13: Ajuste modelo AR Δ volumen suavizado 40 días

Variable		ar1	ar2	ma1	ma2	ma3	ma4
Monto 5 días	coef	0.0824	0.7235	1.3989	0.6636		
	desv. est.	0.0208	0.0232	0.0270	0.0371		
Δ Monto 40 días	coef	0.8457		-0.4893	-0.0868	-0.0039	0.103
	desv. est.	0.0297		0.0369	0.0289	0.0292	0.025

Cuadro 7: Coeficientes modelos para el Monto negociado

```

Series: M5_train
ARIMA(2,0,2) with zero mean

Coefficients:
      ar1      ar2      ma1      ma2
      0.0824  0.7235  1.3989  0.6636
s.e.  0.0208  0.0232  0.0270  0.0371

sigma^2 = 0.07646; log likelihood = -238.05
AIC=486.09  AICc=486.13  BIC=513.53

Training set error measures:
              ME          RMSE          MAE  MPE  MAPE          MASE          ACF1
Training set 0.001028969  0.2762107  0.2175761  4.150759  119.268  0.9012142 -0.1181424
> coeftest(M5mod)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.082419   0.020791   3.9642 7.365e-05 ***
ar2  0.723501   0.023240  31.1312 < 2.2e-16 ***
ma1  1.398891   0.027042  51.7307 < 2.2e-16 ***
ma2  0.663641   0.037084  17.8955 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 14: Ajuste modelo ARMA monto suavizado 5 días

```

Series: Mon40_train
ARIMA(1,0,4) with zero mean

Coefficients:
      ar1      ma1      ma2      ma3      ma4
      0.8457  -0.4894  -0.0868  -0.0040  0.103
s.e.   0.0297  0.0369  0.0290  0.0292  0.025

sigma^2 = 0.004256: log likelihood = 2297.38
AIC=-4582.76  AICC=-4582.71  BIC=-4549.95

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.0001596994  0.06514592  0.05131415  -761.092  1614.95  0.8380606  0.0003126375
> coeftest(arm40)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.845744  0.029690  28.4859 < 2.2e-16 ***
ma1 -0.489386  0.036883 -13.2684 < 2.2e-16 ***
ma2 -0.086842  0.028951  -2.9996  0.002703 **
ma3 -0.004042  0.029181  -0.1385  0.889835
ma4  0.103019  0.024992  4.1221  3.754e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 15: Ajuste modelo ARMA Δ monto suavizado 40 días

9.2. Modelos autoregresivos con variables auxiliares

En este anexo se encuentran las salidas de R que muestran el ajuste de los modelos ARMA con variables auxiliares (ARMAX) para cada una de las variables objetivo, así como las tablas a las que se hace referencia en el documento.

- Para Volumen suavizado 5 días:

$$Y_t = 0,4920Y_{t-1} + 0,5946Y_{t-2} - 0,1332Y_{t-3} + 0,1247Y_{t-4} - 0,2662Y_{t-5} + 0,6610\varepsilon_{t-1} - 0,1220 * TRM_t - 0,0595 * FRB_t + 0,2783 * IPC_t - 0,0725 * Exportados_t - 0,0228 * VIX_t + 0,7306 * Monto_5d_t + \varepsilon_t$$

- Para Monto suavizado 5 días

$$Y_t = -0,2598Y_{t-1} + 0,6616Y_{t-2} + 0,5765Y_{t-3} - 0,0770Y_{t-4} - 0,2665Y_{t-5} + 1,5935\varepsilon_{t-1} + 0,9783\varepsilon_{t-2} + 0,1036 * FRB - 0,1649 * TPM_t + 0,0108 * VIX_t + 0,9743 * Vol_5d_t + \varepsilon_t$$

- Para Volumen suavizado 40 días

$$\Delta Y_t = 0,2383\Delta Y_{t-1} + 0,1054\Delta Y_{t-2} + 0,0656\Delta Y_{t-3} + 0,0993\Delta Y_{t-4} + 0,0986\Delta Y_{t-5} + 0,7136 * Monto_40d_t + \varepsilon_t$$

- Para Monto suavizado 40 días

$$\Delta Y_t = 0,8556\Delta Y_{t-1} - 0,6621\varepsilon_{t-1} - 0,0616\varepsilon_{t-2} + 1,0519 * Vol_40d_t + \varepsilon_t$$

```

Series: v5_train
Regression with ARIMA(5,0,1) errors

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1      TRM      FRB_rate      IPC_prc
s.e.      0.4920  0.5946 -0.1332  0.1247 -0.2662  0.6610 -0.1220 -0.0595  0.2783
      Exportado_MILLONESUSD      CBOE_VIX      Monto_5d
s.e.      -0.0725 -0.0228  0.7306
              0.0192  0.0098  0.0112

sigma^2 = 0.01595: log likelihood = 1165.52
AIC=-2305.04 AICc=-2304.83 BIC=-2233.7

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.001158188 0.1258495 0.09793785 595.0145 666.603 0.4686642 0.05822847
> coeftest(v5mod)

z test of coefficients:

              Estimate Std. Error z value Pr(>|z|)
ar1           0.4919813  0.0353610  13.9131 < 2.2e-16 ***
ar2           0.5946479  0.0400893  14.8331 < 2.2e-16 ***
ar3          -0.1331918  0.0290506  -4.5848 4.544e-06 ***
ar4           0.1246709  0.0273092   4.5652 4.991e-06 ***
ar5          -0.2661776  0.0233680 -11.3907 < 2.2e-16 ***
ma1           0.6610218  0.0297272  22.2363 < 2.2e-16 ***
TRM          -0.1220227  0.0281551  -4.3340 1.465e-05 ***
FRB_rate     -0.0595304  0.0294171  -2.0237 0.0430044 *
IPC_prc       0.2782887  0.0324982   8.5632 < 2.2e-16 ***
Exportado_MILLONESUSD -0.0725183  0.0192459  -3.7680 0.0001646 ***
CBOE_VIX     -0.0227828  0.0098019  -2.3243 0.0201088 *
Monto_5d      0.7305520  0.0112454  64.9645 < 2.2e-16 ***
---

```

Figura 16: Ajuste modelo ARMAX volumen suavizado 5 días

```

Series: M5_train
Regression with ARIMA(5,0,2) errors

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1      ma2      FRB_rate      TPM_BR
s.e.      0.0234  0.0244  0.0254  0.0241  0.0238  0.0080  0.0071  0.0288  0.0349
      CBOE_VIX      vol_5d
s.e.      0.0108  0.9743
              0.0036  0.0146

sigma^2 = 0.01799: log likelihood = 1055.21
AIC=-2086.42 AICc=-2086.24 BIC=-2020.57

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -0.0006121493 0.133706 0.106133 4.171618 55.38007 0.4396096 -0.01867129
> coeftest(M5mod)

z test of coefficients:

              Estimate Std. Error z value Pr(>|z|)
ar1          -0.2598477  0.0234134 -11.0982 < 2.2e-16 ***
ar2           0.6615855  0.0243951  27.1196 < 2.2e-16 ***
ar3           0.5765332  0.0253857  22.7109 < 2.2e-16 ***
ar4          -0.0769794  0.0241112  -3.1927 0.0014096 **
ar5          -0.2664897  0.0237585 -11.2166 < 2.2e-16 ***
ma1           1.5934831  0.0080389  198.2225 < 2.2e-16 ***
ma2           0.9782909  0.0070674  138.4234 < 2.2e-16 ***
FRB_rate      0.1035895  0.0287514   3.6029 0.0003146 ***
TPM_BR       -0.1648546  0.0348549  -4.7297 2.248e-06 ***
CBOE_VIX      0.0108147  0.0036488   2.9639 0.0030381 **
vol_5d        0.9742910  0.0146306  66.5929 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 17: Ajuste modelo ARMAX monto suavizado 5 días

```

Series: vol40_train
Regression with ARIMA(5,0,0) errors

Coefficients:
      ar1      ar2      ar3      ar4      ar5      xreg
s.e.  0.0241  0.0244  0.0245  0.0244  0.0239  0.0102

sigma^2 = 0.0007288:  log likelihood = 3842.96
AIC=-7671.92  AICC=-7671.85  BIC=-7633.64

Training set error measures:
              ME          RMSE          MAE  MPE  MAPE
Training set -0.0002056862  0.02695056  0.02092875 -Inf  Inf
              MASE          ACF1
Training set  0.4267916 -0.00305498
> coeftest(v40mod)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.238282   0.024076  9.8969 < 2.2e-16 ***
ar2  0.105359   0.024432  4.3123 1.616e-05 ***
ar3  0.065639   0.024469  2.6826 0.007306 **
ar4  0.099286   0.024409  4.0676 4.751e-05 ***
ar5  0.098557   0.023858  4.1310 3.612e-05 ***
xreg 0.713561   0.010209 69.8953 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 18: Ajuste modelo ARX Δ volumen suavizado 40 días

```

Series: Mon40_train
Regression with ARIMA(1,0,2) errors

Coefficients:
      ar1      ma1      ma2      xreg
      0.8556 -0.6621 -0.0616  1.0519
s.e.  0.0429  0.0497  0.0288  0.0139

sigma^2 = 0.001066:  log likelihood = 3509.33
AIC=-7008.67  AICC=-7008.63  BIC=-6981.33

Training set error measures:
              ME          RMSE          MAE  MPE  MAPE
Training set 0.0004000252  0.03260919  0.02559698 2121.204 2342.831
              MASE          ACF1
Training set  0.4180487 -0.0005028757
> coeftest(m40mod)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.855567   0.042929 19.9300 < 2e-16 ***
ma1 -0.662079   0.049674 -13.3286 < 2e-16 ***
ma2 -0.061591   0.028760 -2.1416 0.03223 *
xreg 1.051890   0.013867 75.8575 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 19: Ajuste modelo ARMAX Δ monto suavizado 40 días

9.3. Optimización

	LSTM	TCN	XGBOOST
Tasa aprendizaje	(1e-06, 1e-02)	(1e-06, 1e-02)	(1e-06, 1e-02)
Dropout	(0.1, 0.4)	(0.1, 0.4)	(0.1, 0.4)
Tamaño lote	(20,40)	(20,40)	
Tamaño secuencia	(15,25)		
Nº nodos por capa profundidad	(12,65)		(2,10)
Tamaño kernel		(2,5)	
Filtros		(1,3)	
Dilatacion		(0,3)	

Cuadro 8: Espacio de búsqueda hiperparámetros

9.4. Test Diebold-Mariano

En las siguientes tablas se pueden ver los estadísticos de prueba y los p-valores correspondientes a la aplicación del test modificado de Diebold-Mariano tomado de [30]. Siendo e_{1t} y e_{2t} los errores de pronóstico del modelo_1 y modelo_2, y $g(\cdot)$ una función de pérdida, sea $d_t = g(e_{1t}) - g(e_{2t})$. Se usará la prueba a una cola dejando la hipótesis nula como $H_0 : E(d_t) = 0 \forall t$ vs la hipótesis alterna $H_1 : E(d_t) < 0$.

Entiéndase modelo_1 como los modelos que están las columnas y modelo_2 los modelos que están listados en las filas, así para el caso de tabla 9 donde se compara el modelo ARMA con el ARMAX se dice que con un p-valor de 1 no existe suficiente evidencia estadística para rechazar la hipótesis de que el modelo ARMAX (modelo_2) tiene el mismo error de pronóstico que el modelo ARMA (modelo_1).

		ARMA	ARMAX	LSTM	TCN
ARMAX	Est.	6.0928			
	P. Valor	1.0000			
LSTM	Est.	5.2212	-0.3261		
	P. Valor	1.0000	0.3734		
TCN	Est.	-6.1985	-8.1456	-7.3286	
	P. Valor	0.0000	0.0000	0.0000	
XGBOOST	Est.	6.0059	3.4502	2.9080	7.6636
	P. Valor	1.0000	0.9991	0.9965	1.0000

Cuadro 9: Test Diebold-Mariano volumen transacciones 5 días

		ARMA	ARMAX	LSTM	TCN
ARMAX	Est.	5.1733			
	P. Valor	1.0000			
LSTM	Est.	5.0755	-1.2553		
	P. Valor	1.0000	0.1095		
TCN	Est.	-5.8861	-6.4658	-6.1410	
	P. Valor	0.0000	0.0000	0.0000	
XGBOOST	Est.	5.3502	0.2743	2.4159	6.4684
	P. Valor	1.0000	0.6071	0.9890	1.0000

Cuadro 10: Test Diebold-Mariano monto negociado 5 días

		ARMA	ARMAX	LSTM	TCN
ARMAX	Est.	1.3509			
	P. Valor	0.9066			
LSTM	Est.	3.0093	1.5369		
	P. Valor	0.9974	0.9326		
TCN	Est.	-0.1501	-3.5194	-3.1907	
	P. Valor	0.4408	0.0007	0.0017	
XGBOOST	Est.	0.6776	-1.2841	-3.2444	0.6159
	P. Valor	0.7484	0.1045	0.0014	0.7287

Cuadro 11: Test Diebold-Mariano volumen transacciones 40 días

		ARMA	ARMAX	LSTM	TCN
ARMAX	Est.	2.5874			
	P. Valor	0.9926			
LSTM	Est.	5.5550	14.1040		
	P. Valor	1.0000	1.0000		
TCN	Est.	5.1734	0.5568	-5.0722	
	P. Valor	1.0000	0.7091	0.0000	
XGBOOST	Est.	5.4904	13.2180	-2.6764	4.8341
	P. Valor	1.0000	1.0000	0.0060	1.0000

Cuadro 12: Test Diebold-Mariano monto negociado 40 días