

Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales

Predicting the Benchmark Banking Indicator (IBR) Time Series through Neural Networks

Germán Enrique Coy Mondragón^{ab}, Óscar Granados^{ac}, Olmer Garcia-Bedoya^{ad}

^a Universidad de Bogotá Jorge Tadeo Lozano, Colombia

^b germane.coymondragon@utadeo.edu.co

^c <https://orcid.org/0000-0002-4992-8972>

^d <https://orcid.org/0000-0002-6964-3034>

RESUMEN

En años recientes, la predicción del comportamiento del Indicador Bancario de Referencia (IBR) se ha hecho relevante debido a su importancia en el mercado monetario colombiano. El propósito de este trabajo es demostrar la eficiencia de las redes LSTM en la generación de predicciones de series de tiempo —a través de su memoria a corto y largo plazo— que sean comparables con el modelo predictivo ARIMA para estudios econométricos. Se analizó la incidencia de la tasa representativa de mercado (TRM) y la tasa de los bonos de la deuda pública (TES) a 10 años, comparando ambos indicadores con el IBR. Con lo anterior, se buscó determinar la correlación existente entre estas variables mediante el método de Pearson. Finalmente, la eficiencia del modelo fue evaluada con el error cuadrático medio (RMSE), utilizando una red LSTM multivariable con tres entradas (IBR, TES y TRM) y una salida.

Palabras clave: series de tiempo, aprendizaje automático, redes LSTM, ARIMA, coeficiente de correlación de Pearson.

Citation: Coy-Mondragón, G. E., Granados, O. y Garcia-Bedoya, O. (2021).

Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales. *Mutis*, 11(1), 65-76.

<https://doi.org/10.21789/22561498.1748>

Recibido: Enero 18, 2021. **Aceptado:** Marzo 30, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.

ABSTRACT

In recent years, predicting the behavior of the Benchmark Banking Indicator (IBR) has become relevant due to its importance in the Colombian money market. The purpose of this paper is to demonstrate the efficiency of LSTM networks for generating predictions of time series —through their long and short-term memory— that are comparable with the ARIMA predictive model for econometric studies. The incidence of the representative market rate (TRM) and

the rate of 10-year public debt bonds (TES) was analyzed and compared to the IBR, seeking to determine its correlation through the Pearson method. Finally, the model efficiency was evaluated with the mean square error (RMSE), using a multivariable LSTM network with three inputs (IBR, TES, and TRM) and one output.

Keywords: Time series, machine learning, LSTM network, Root Mean Square Error (RMSE), ARIMA, Pearson Correlation Coefficient.

INTRODUCCIÓN

El Banco de la República y el Ministerio de Hacienda y Crédito Público de Colombia, con el propósito de tener un indicador confiable que reflejara la liquidez del mercado monetario, crearon en 2008 el Indicador Bancario de Referencia (IBR). Este indicador relaciona el precio al que los bancos están dispuestos a ofrecer o captar recursos en el mercado interbancario, lo que convierte al IBR en una referencia indispensable para las operaciones financieras de corto plazo. En este mecanismo se establecen cuatro plazos de cotización (*overnight*, un mes, tres meses y seis meses) y a partir de él se calcula la mediana de las cotizaciones de los participantes que se publica en los sistemas de información del Banco de la República. Varios mercados utilizan indicadores semejantes que se convirtieron en referentes para el IBR, principalmente, el London InterBank Offered Rate (LIBOR) de Londres, el Euro Interbank Offered Rate (EURIBOR) de la zona Euro y la tasa de interés interbancaria de equilibrio (TIIE) de México, por listar algunos (Abrantes-Metz *et al.*, 2001; Duffie & Stein, 2015; Jamshidian, 1997). Estos indicadores utilizan las tasas de referencia interbancaria, por lo que varios estudios han buscado determinar su pronóstico a través de técnicas de aprendizaje automático y, más recientemente, de aprendizaje profundo.

En Colombia se registran pocos trabajos similares con esta clase de técnicas. Por ello, entendiendo la relevancia de estos indicadores para los mercados financieros, surge la necesidad de desarrollar un modelo de predicción del comportamiento del IBR. Diversos estudios han utilizado modelos de series temporales, con la ayuda de los principios del aprendizaje automático y el uso de técnicas como la red LSTM, para hallar la predicción de la tendencia en diferentes casos de estudio, arrojando importantes resultados. En los últimos años, se ha utilizado la metodología LSTM para predecir y clasificar texto con la ayuda de redes neuronales recurrentes (RNN), entre ellas, las redes LSTM (Liu *et al.*, 2016), así como para predecir las

cargas de trabajo (Kumar *et al.*, 2018), la dependencia a largo plazo de series temporales con datos financieros (Zhang *et al.*, 2019), la detección temprana de un shock séptico con el objetivo de reducir la mortalidad en pacientes (Fagerström *et al.*, 2019), los precios de las acciones en la apertura del mercado bursátil (Qui *et al.*, 2020), la trayectoria de un vehículo con datos de condición reales (Bükk & Johansson, 2020) o los lugares para la exploración minera (Li *et al.*, 2020), por citar algunos ejemplos.

Algunos estudios usan las metodologías de aprendizaje automático para predecir indicadores financieros y económicos, empleando el algoritmo genético puro (GA) y el algoritmo genético integrado con la red neural recurrente (GA + RNN) para predecir la tasa LIBOR en USD (Tan, 2019). Así mismo, existen propuestas que buscan predecir a través de redes neuronales el cambio diario del precio de la tasa de interés LIBOR y el índice bursátil estadounidense Standard and Poor's 500 (S&P 500) (Dmitrieva *et al.*, 2002), o pronosticar las acciones de mayor bursatilidad en Colombia (Ecopetrol y Preferencial Bancolombia) que se encuentran cotizadas en el mercado local y en el estadounidense, a través del esquema de American Depositary Receipt (ADR) (Villada *et al.*, 2012), por ejemplo. Otros trabajos han usado esta metodología para pronosticar variables económicas como la inflación, para el caso de México, a través de una red neural artificial (RNA) con un enfoque estacional y un proceso AR (Robles, 2019); en Costa Rica, al comparar el rendimiento de algunas técnicas de *machine learning* (K-Nearest Neighbours [KNN], bosques aleatorios, *extreme gradient boosting* y un modelo de tipo *long short-term memory* [LSTM]) (Rodríguez-Vargas, 2020); o en Chile para pronosticar la inflación subyacente del IPC con métodos de *machine learning* y *big data* (Leal *et al.*, 2020). Finalmente, existen algunos trabajos que hacen análisis mediante técnicas de *machine learning* como las redes LSTM y ARIMA (*autoregressive integrated moving average*) para predecir el precio de cierre de las criptomonedas (Castillo-Mur, 2020).

Teniendo en cuenta lo anterior, ¿cómo podría aplicarse esta metodología a un indicador interbancario?, ¿cómo integrarla en la interacción con otras variables financieras? A partir de estos interrogantes, la presente investigación busca contribuir en la definición y generación de una predicción confiable de la serie temporal del IBR y analizar en un ambiente comparativo la eficiencia de redes LSTM tanto supervisadas como no supervisadas. Adicionalmente, esta investigación busca establecer la confiabilidad y las limitantes del uso de herramientas de aprendizaje automático en variables financieras en las que habitualmente los análisis econométricos han sido ampliamente usados.

Para lograr los objetivos propuestos, este trabajo usa la metodología CRISP-DM, la cual permite realizar proyectos mediante una estructura de tareas por nivel o procesos (Wirth *et al.*, 2000). Esta metodología se fundamenta, primero, en la comprensión del negocio —definida en esta introducción— y se apoya en la sección de preliminares. Además, se apoya en la comprensión y preparación de los datos que se desarrollan en la metodología, junto con el modelo y su implementación. La evaluación de los resultados se integra en la sección de resultados. Finalmente, la evaluación de los resultados y la identificación de la utilidad del modelo se presenta en la sección de discusión.

PRELIMINARES

En esta sección se hace una revisión básica de algunos conceptos que son empleados en el desarrollo del documento y que facilitan el entendimiento de su integración.

Indicador bancario de referencia (IBR)

La investigación se desarrolló con el esquema *overnight* del IBR, el cual se comporta de la siguiente manera, de acuerdo con Cardozo y Murcia (2014):

- Las cotizaciones por parte de las entidades financieras participantes se realizan diariamente entre 10:00 a. m. y 10:15 a. m. Luego, el Banco de la República toma en cuenta solo la última tasa de cotización recibida de cada entidad participante.
- El Banco de la República publica, cada día hábil a las 11:00 a. m., el IBR *overnight*.

- Las transferencias de los recursos interbancarios se deberán realizar antes del cierre de la plataforma de Servicios Electrónicos del Banco de la República (SEBRA).
- El pago o la devolución de los créditos interbancarios entre los participantes se realiza de manera automática a través de la plataforma que el Banco de la República disponga para el efecto.

Coefficiente de correlación

El coeficiente de correlación busca establecer si existe alguna asociación entre dos variables cuantitativas aleatorias que tienden a ser mayores (o menores) a medida que aumentan los valores de la otra variable; su correlación puede determinarse mediante el resultado, que puede fluctuar entre 1 y -1, mientras que su proximidad a 0 puede indicar la no correlación o baja correlación lineal (ec. 1).

$$-1 \leq \rho \leq 1 \quad (\text{Ec. 1})$$

Coefficiente de correlación Pearson

Con el coeficiente de correlación Pearson se puede observar las similitudes ρ entre las variables o los individuos. Así mismo, es posible medir la fuerza o el grado de asociación que posee una distribución normal bivariada conjunta (Restrepo & González, 2007). Este coeficiente se expresa de la siguiente manera (ec. 2):

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (\text{Ec. 2})$$

donde:

- si ρ es positivo la relación es directa entre las variables;
- si ρ es negativo la relación entre las variables es inversa, es decir, cuando una aumenta, la otra disminuye;
- si ρ es mayor de 1 o -1 hay una relación perfecta entre ambas variables;
- si ρ es igual a 0 no hay relación entre las variables; sin embargo, puede existir una relación no lineal.

ARIMA

El modelo ARIMA (*autoregressive integrated moving average*) es un proceso estocástico para el estudio de series temporales en temas econométricos desarrollado por Box y Jenkins. Este modelo se deriva de tres componentes —AR (autorregresivo), I (integrado) y MA (medidas móviles)— y logra interpretar valores en funciones lineales con componentes cíclicos o estacionales. Además, este supone valores futuros combinando valores pasados, si estos son lineales (Pai & Lin, 2005), como se expresa en la ecuación 3. Sin embargo, si el grupo de datos no corresponden a una linealidad o no se adecua a este comportamiento, el modelo ARIMA se ve limitado.

$$\rho = y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (\text{Ec. 3})$$

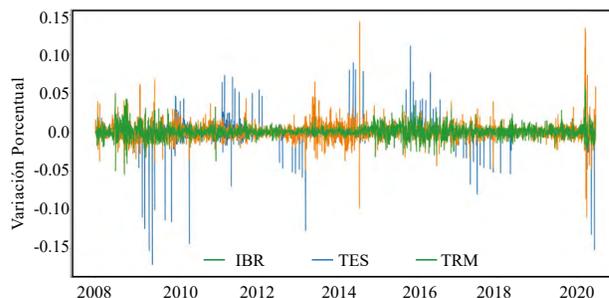
MATERIALES Y MÉTODOS

Preparación de los datos

Los datos de entrada están constituidos por el IBR, los títulos de deuda pública a 10 años (TES 10 años) y la tasa representativa del mercado (TRM) durante el período comprendido entre el 2 de enero de 2008 y el 15 de junio de 2020. Los datos se obtuvieron del conjunto de datos públicos del Banco de la República de Colombia y fueron ajustados para ser cargados en Python.

El conjunto de datos del plazo *overnight* del IBR está dividido en 21 columnas (información proporcionada por los bancos participantes para el cálculo del indicador y el cálculo del IBR nominal y efectivo) y se define como un valor relativo. Los datos con los que se comparó el indicador fueron la TRM, expresada en pesos colombianos (COP) por dólar estadounidense (USD), y los rendimientos de los títulos de deuda pública a diez años (TES), expresados en la tasa de rendimiento, que es un valor relativo. Estos últimos se tomaron en el mismo rango de tiempo y, al igual que el IBR, fueron descargados en formato csv, siendo necesario mejorar la calidad de los datos antes de empezar su análisis con el modelo de aprendizaje automático. Para lograr datos óptimos, se determinó la tasa de variación de cada uno de los conjuntos de datos comparados con el fin de obtener datos semejantes y así poder analizar de forma más clara los cambios que experimentan los datos a través del tiempo (figura 1).

Figura 1. Tasa de variación



Fuente: elaboración propia.

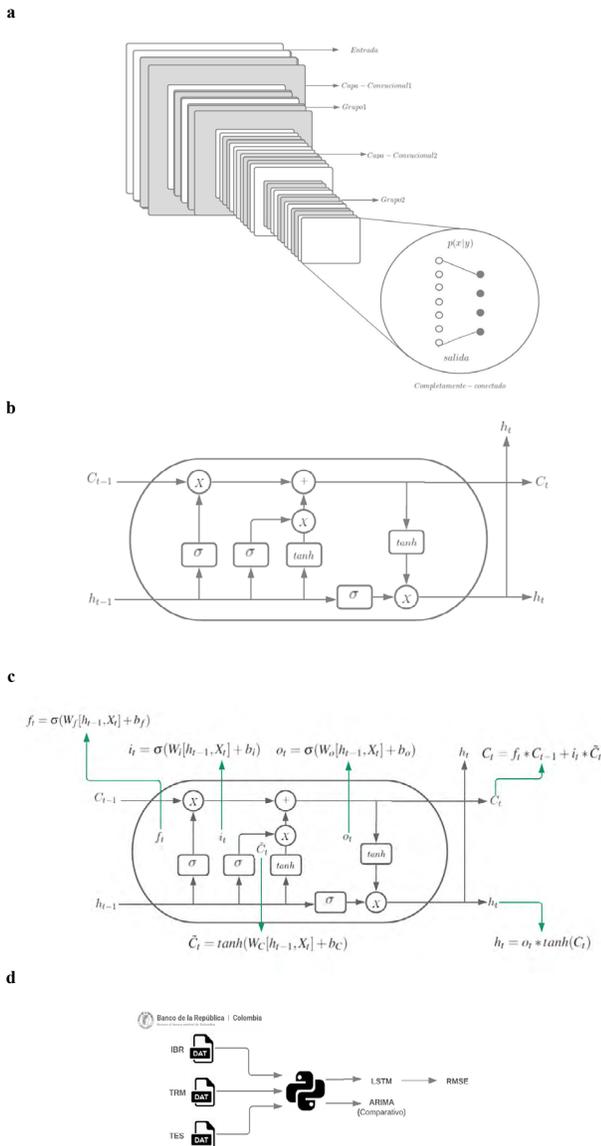
Modelo

El método para la predicción del IBR está constituido por una red neuronal recurrente LSTM (figura 2a) que aprende de los datos históricos a largo plazo. Este método, utilizado en aprendizaje profundo (*deep learning*), aprende los valores futuros de una serie temporal en función de los valores anteriores y, como toda RNN, permite conexiones arbitrarias entre las neuronas, generando temporalidad y, desde luego, memoria.

La red LSTM tiene una arquitectura recurrente con conexiones que retroalimentan las capas de la red, permitiendo analizar los mismos problemas que una máquina de Turing. Una red neuronal recurrente tradicional está compuesta por una sola capa que calcula el peso que van a tener los datos históricos. En las redes LSTM del sistema de persistencia se definen cuatro puertas, o unidades, que interactúan entre sí (figura 2b). Por lo tanto, las redes LSTM solucionan el problema de largo plazo con el que se encuentran las redes neuronales recurrentes tradicionales, logrando resolver el

problema que surge cuando se necesitan datos de muchos ciclos atrás para realizar la predicción del ciclo siguiente (Pastor-Cobo & Torre-Barrio, 2019).

Figura 2. Modelo LSTM



a. Estructura LSTM implementada; b. Primer modelo LSTM; c. Segundo modelo LSTM; d. Proceso del modelo.

Fuente: elaboración propia.

Este algoritmo LSTM utiliza una de las formas de RNN más comunes y es capaz de procesar una secuencia de longitudes arbitrarias, aplicando recursivamente una función de transición a su vector de estado oculto

interno en las secuencias de entrada (Liu *et al.*, 2016). Además, este recurso evita problemas de dependencia a largo plazo para la predicción de series de tiempo. Según Hochreiter y Schmidhuber (citados en Qui *et al.*, 2020), este modelo consiste en un conjunto único de células de memoria que reemplazan las neuronas de la capa oculta del RNN, y su clave es el estado de las celdas de memoria que filtran información a través de la estructura de la puerta. Matemáticamente, una celda de la red LSTM se representa de la siguiente manera (figura 2c).

- Entrada

X_t = valor de la secuencia en el instante t

h_{t-1} = salida de la unidad LSTM en el paso anterior

C_{t-1} = estado de la unidad LSTM

- Salida

Se conforma por un nuevo estado o una neurona autorrecurrente, que es la combinación entre C_{t-1} y el candidato a nuevo estado \tilde{C}_{t-1} , donde podemos ver f_t y i_t como el grado de influencia de C_{t-1} y \tilde{C}_t .

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (\text{Ec. 4})$$

La salida de la unidad o celda sería:

$$h_t = o_t * \tanh(C_t) \quad (\text{Ec. 5})$$

Para hallar el nuevo candidato se utiliza la siguiente ecuación:

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, X_t] + b_C) \quad (\text{Ec. 6})$$

donde $W_C[h_{t-1}, X_t]$ representa las entradas y los datos recurrentes, y $[h_{t-1}, X_t]$ concatena los vectores h_{t-1}, X_t y b_C .

Con el nuevo candidato hallado se calcula un nuevo estado:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (\text{Ec. 7})$$

Ahora, se calcula el f_t *forget gate*, o puerta de olvido,

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (\text{Ec. 8})$$

y la i_t *external input gate*, o puerta de entrada externa,

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \quad (\text{Ec. 9})$$

Para terminar, se calcula otra vez el nuevo estado con la ecuación 7, se calcula la salida de la red LSTM con la ecuación 5, y se determina O_t *output gate*, o puerta de salida.

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad (\text{Ec. 10})$$

Finalmente, para validar los datos obtenidos primero se compararán los resultados de la red LSTM con los de ARIMA, a fin de determinar su calidad en la predicción. Seguidamente, para determinar el error del modelo se evalúa el error cuadrático medio (RMSE, en inglés) con una red LSTM multivariable (figura 2d).

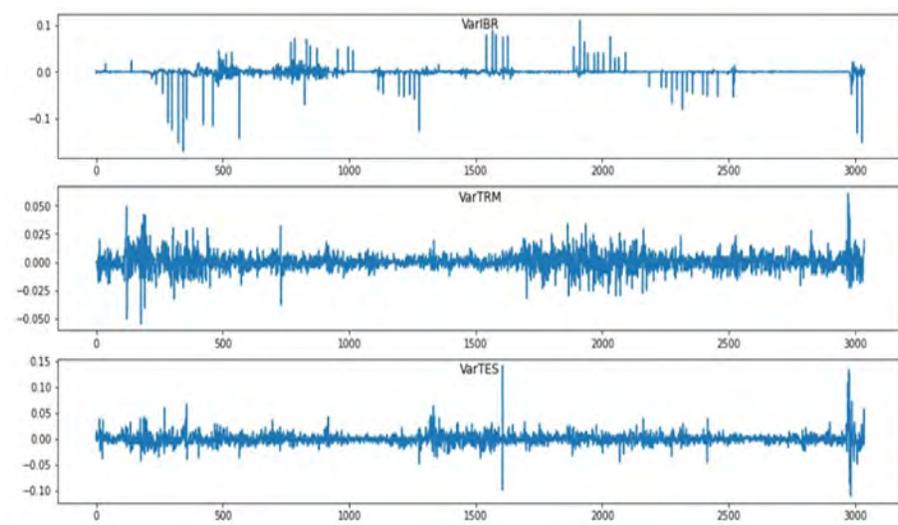
Implementación del modelo

Predicción del IBR, TES y TRM

En esta primera etapa de la investigación se determinó la predicción independiente de cada una de las series temporales definidas en el mismo rango de tiempo (figura 3), haciendo uso de redes LSTM con 50 neuronas en la primera capa oculta y una neurona en la capa de salida. Esta última fue ajustada a 10 épocas con un tamaño de 20 lotes.

Posteriormente, se evaluó la predicción de estas mismas variables con el modelo ARIMA, comparando sus resultados con los hallazgos obtenidos de la predicción con la red LSTM. Finalmente, se calcularon algunos comportamientos estadísticos de las variables y se determinó su correlación.

Figura 3. Señales de los indicadores usados IBR, TES y TRM



Fuente: elaboración propia.

Cálculo del error de la red LSTM

En esta segunda etapa de la investigación, tras comprobar la eficiencia y calidad de la predicción del modelo de aprendizaje automático LSTM frente al ARIMA, se procedió a comprobar la validez del modelo propuesto con ayuda de una red LSTM múltiple con tres entradas (ver figura 7) y una salida, donde se evaluó el modelo con ayuda del RMSE (ecuación 11) y el error absoluto medio (MAE) (ec. 12).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (\text{Ec. 11})$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (\text{Ec. 12})$$

Mediante el uso del error cuadrático medio, se evalúa la diferencia entre los datos reales y los pronosticados, calculando la media de los cuadrados del error. De otro lado, con el uso de MAE se puede limitar el error, donde al elevar al cuadrado la diferencia solo se observa los errores más altos, afectando así la calidad del resultado final.

Inicialmente, se normalizó las series temporales, o señales de entrada, y se convirtió el problema a uno de tipo aprendizaje supervisado. La importancia de hacer este cambio radica en la necesidad de contar con patrones de entrada (x) y salida (y) para que el algoritmo pueda aprender a predecir el patrón salida con los patrones de entrada, convirtiendo los DataFrame con $t-1$ como x y t como y . Al ejecutar el código podemos ver las tres series temporales (señales) de entrada y la variable $var1(t)$, que corresponde a la predicción de salida (figura 4).

Figura 4. Conjunto de datos transformados

	var1(t-1)	var2(t-1)	var3(t-1)	var1(t)
1	0.607475	0.489499	0.438874	0.615705
2	0.615705	0.428184	0.500116	0.598145
3	0.598145	0.508808	0.442443	0.612714
4	0.612714	0.517276	0.419458	0.611586
5	0.611586	0.150407	0.464547	0.609342

(3036, 4)

Fuente: elaboración propia.

Posteriormente, se definió y se ajustó el modelo, asignando 50 neuronas para la primera capa oculta y una neurona para la capa de salida, junto con 50 épocas de entrenamiento y un tamaño de 20 lotes. Finalmente, se ejecuta el modelo con aproximadamente 200 horas de datos para entrenamiento y 2.800 horas para pruebas, tanto para las variables de entrada como para las de salida, realizando seguimiento al entrenamiento y a la pérdida de pruebas durante este.

RESULTADOS

Con el fin de identificar la incidencia entre los conjuntos de datos del IBR y su comparación con la TRM y la tasa de rendimiento de los TES a 10 años, se encontraron los siguientes resultados.

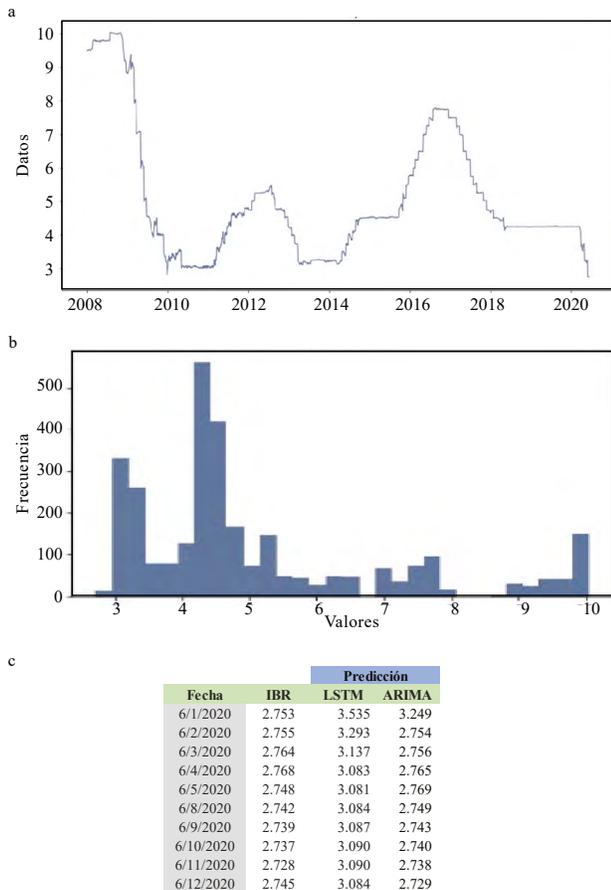
Análisis del IBR

El IBR surgió como respuesta a las anomalías crediticias desencadenadas por el DTF y con el objetivo de elevar estándares de credibilidad y transparencia. Desde sus inicios, en 2008, este indicador presentó algunas fluctuaciones crecientes y decrecientes, siendo las más significativas aquellas ocurridas en los periodos 2008-2010 y 2016-2018. En 2020, desde marzo, se evidencia que este indicador decreció hasta 2,742 %, como resultado del impacto de la pandemia por COVID-19 (figura 5a).

Según la frecuencia de los datos, se determinó que la mayoría de estos se encuentran en un rango de 4-5 % y fluctúa entre 2,728 y 10,044 %, con una media = 5,0661, mediana = 4,499, desviación estándar = 1,9098 y varianza = 3,6474 (figura 5b). Adicionalmente, se comparó la eficiencia de los modelos predictivos Red LSTM con los modelos ARIMA (figura 5c). De acuerdo con los datos generados, la confiabilidad de estos dos modelos predictivos es alta, aun cuando se tomó una

muestra de tiempo en la cual la fluctuación del indicador se pudo ver afectada por el impacto de la pandemia de COVID-19.

Figura 5. Resultados IBR



Fuente: elaboración propia.

Análisis de los títulos de deuda pública (TES 10 años)

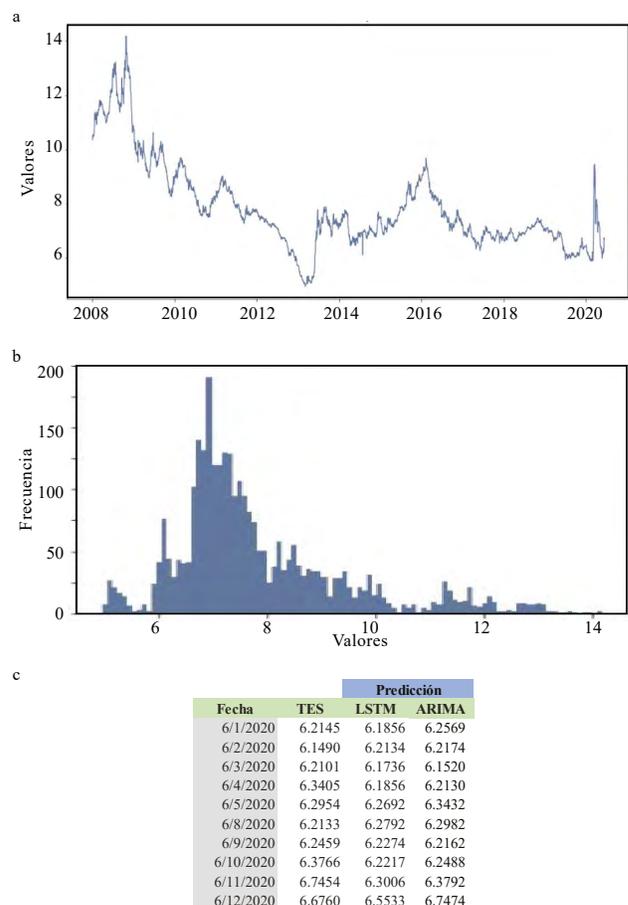
La importancia de comparar la tasa de rendimiento de los TES con el IBR radica en que los TES son el título de mayor inversión en el mercado de capitales, siendo los bancos quienes dan liquidez a los bonos a través del programa de creadores de mercado. Por tal razón, la emisión de bonos se relaciona en cierto modo con el IBR; además de estar estructurados sobre un cupón que corresponde a una tasa de interés.

Durante el período 2009-2015, el comportamiento de los TES presentó una caída en el rendimiento. En 2020, motivado por la pandemia, se evidencia un incremento en el rendimiento hasta el 24 de marzo, momento

en el cual nuevamente comenzó a descender, hasta alcanzar un mínimo de 5,985 % (figura 6a).

La frecuencia de los datos determina que la mayoría de los TES se encuentra en un rango entre 6-8 %, fluctuando entre 4,94651388919636 y 14,18142693689 % (figura 6b), con una media = 7,8261, mediana = 7,3593, desviación estándar = 1,6028 y varianza = 2,5691. Al igual que se hizo con los datos del IBR, se compararon los datos de los TES con los modelos predictivos Red LSTM y ARIMA, generando resultados que confirman su confiabilidad tras las predicciones con ambos modelos (figura 6c).

Figura 6. Resultados TES



Fuente: elaboración propia.

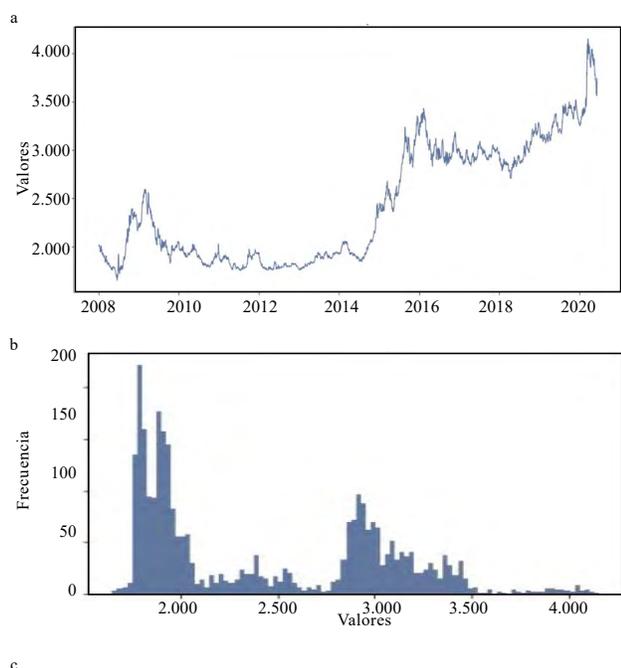
Análisis de la tasa representativa del mercado (TRM)

La TRM ha fluctuado por diferentes motivos, tanto globales como de la economía colombiana. El comportamiento de la TRM durante el período analizado

presenta una tendencia alcista, aunque cabe destacar que este crecimiento se presenta de manera más visible desde finales de 2015, como se muestra en la figura 7a. Además, es preciso mencionar que, por motivos de la pandemia, en el primer semestre de 2020 se llegó a un valor máximo de la TRM de 4.153,91 pesos colombianos (COP) por cada dólar estadounidense (USD).

Por otro lado, la frecuencia de los datos de la TRM determina que la mayoría de estos se encuentran en un rango entre COP 1.800 y 2.100, fluctuando entre COP 1.625,41 y 4.153,9, con una media = 2.423,7979 y una mediana = 2.158,5 (figura 7b). Finalmente, al igual que con los datos del IBR y los TES, se compararon los datos de la TRM con los modelos predictivos Red LSTM y ARIMA, generando una confiabilidad aceptable en ambos casos (figura 7c).

Figura 7. Resultados TRM



Fuente: elaboración propia.

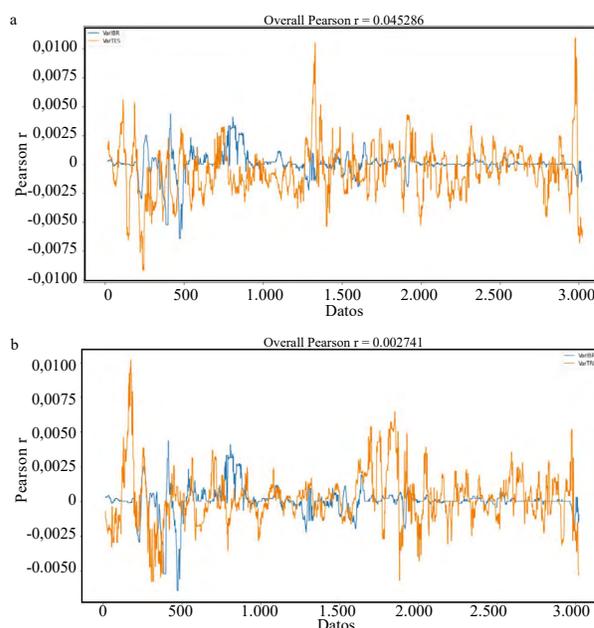
Análisis de correlación

Se determinó el coeficiente de correlación con el método Pearson y se analizaron tres variables cuantitativas (IBR, TES y TRM), las cuales fueron transformadas previamente a tasa de variación y permitieron obtener los siguientes resultados (figura 8):

- Análisis IBR/TES: $\rho = 0,045286$
- Análisis IBR/TRM: $\rho = 0,002741$

En los resultados se evidencia que no se cumple el supuesto de normalidad bivariada conjunta debido a que la variable ancho no se ajusta a la distribución gaussiana (o normal). Así, es posible que exista una relación no lineal entre las variables evaluadas debido a su cercanía a 0.

Figura 8. Coeficiente de correlación Pearson: a. Coeficiente para IBR/TES, b. Coeficiente para IBR/TRM



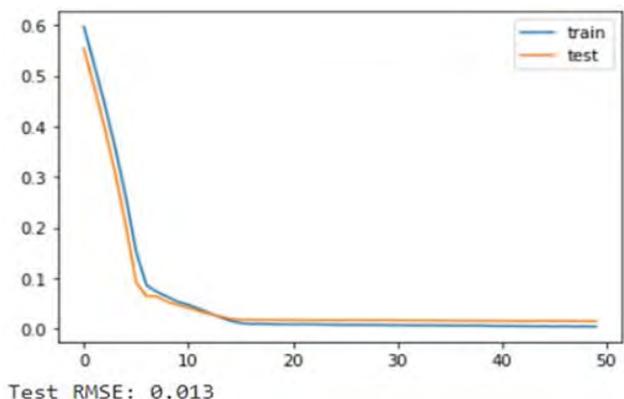
Fuente: elaboración propia.

Análisis del error del modelo de la red LSTM supervisada

Con el modelo ajustado y los pronósticos del conjunto de datos de prueba, se calculó la puntuación del error del modelo empleando el RMSE. El resultado obtenido fue de 0,013 (figura 9), lo que confirma la eficiencia de la red LSTM en la predicción. Para este caso, se utilizó

la versión eficiente de Adam del descenso de gradiente estocástico y la función de pérdida de MAE.

Figura 9. Gráfico de pérdida en el entrenamiento y conjunto de datos de prueba



Fuente: elaboración propia.

DISCUSIÓN Y CONCLUSIÓN

Teniendo en cuenta el objetivo principal de la investigación, se puede afirmar que la fiabilidad de los datos obtenidos con el modelo propuesto es aceptable (red LSTM). Los resultados contemplan un margen de error bajo entre los datos reales y predichos, observando una alta calidad en los datos generados. Desde el punto de vista del despliegue de la metodología CRISP-DM, se puede determinar que es factible llevar a cabo investigaciones tanto en ambientes financieros y económicos —en los que habitualmente se usan metodologías econométricas— como en otras áreas, permitiendo producir resultados de calidad. Si bien algunos pasos de la metodología son de alta complejidad y conllevan mucho tiempo, como la extracción y limpieza de datos o la elección del modelo, esta brinda confiabilidad en el resultado final.

Con el desarrollo paso a paso de la metodología CRISP-DM y su ciclo de vida para proyectos de minería de datos, se pudo evidenciar todo el procedimiento y el tratamiento de los datos, desde su comprensión hasta el despliegue, incluyendo fases vitales como el

modelado de datos, en el que se determinó la predicción del IBR con ayuda de una red LSTM multivariable supervisada, así como su RMSE.

Después de comprobar la eficiencia en la predicción de la temporalidad del indicador a través del método tradicionalmente empleado para la predicción de series temporales ARIMA y someter los datos a un sistema de red LSTM multivariable con tres entradas y una salida, determinando además su RMSE, concluimos que los resultados obtenidos con la red LSTM y su procesamiento interno de cuatro puertas (incluyendo una puerta de olvido) robustece la predicción, haciendo que esta sea altamente confiable. Estos resultados son semejantes a los datos obtenidos con otro modelo predictivo. Además, el error es bajo al comparar el valor observado con el conocido.

Al fijar los datos de entrada para la red LSTM multivariable, además de determinar el IBR, se buscó identificar la relación de este indicador con los TES y la TRM durante un período de 10 años con tasa de variación, donde igualmente fue posible determinar la eficiencia del modelo para una red LSTM simple, logrando obtener buenos resultados, en comparación con ARIMA. Finalmente se analizó la correlación de estas tres variables (IBR, TES y TRM) a través del método de Pearson, logrando evidenciar una correlación positiva muy baja, lo cual señala la posibilidad de que estas variables se relacionen entre sí de una forma no lineal.

Los resultados obtenidos con respecto al comportamiento de la predicción mediante el uso de la red LSTM y su sistema de puertas (que proporciona una fuente de memoria y de información procesada por unidades o celdas que se adaptan) son claros y abren una puerta para futuras investigaciones con datos financieros y fines predictivos. Al observar la baja correlación entre la IBR, los TES y la TRM, se podría considerar la realización de nuevos estudios con métodos de aprendizaje automático supervisado o no supervisado, con el fin de lograr una mejor comprensión del comportamiento de estas variables y de los efectos de su posible estacionalidad.

REFERENCIAS

- Abrantes-Metz, R. M., Villas-Boas, S. B., & Judge, G. (2011). Tracking the LIBOR rate. *Applied Economics Letters*, 18(10), 893-899. <https://doi.org/10.1080/13504851.2010.515197>
- Bükk, A., & Johansson, R. (2020). *Vehicle trajectory prediction using recurrent LSTM neural networks* (tesis de maestría, Chalmers University of Technology). https://odr.chalmers.se/bitstream/20.500.12380/300726/1/B%C3%BCKk_Johansson_2020.pdf
- Cardozo, P., & Murcia, A. (2014). *El IBR y la DTF como indicadores de referencia*. Reportes del Emisor n.º 187. Banco de la República de Colombia.
- Castillo-Mur, A. (2020). *Las criptomonedas en el sistema financiero internacional. Un análisis predictivo mediante técnicas de machine learning (LSTM)* (tesis de pregrado Universidad de Barcelona/Universidad Politécnica de Catalunya). Dipòsit Digital de la Universitat de Barcelona. <http://diposit.ub.edu/dspace/handle/2445/171827>
- Dmitrieva, L., Kuperin, Y., & Soroka, I. (2002). Neural network prediction of short-term dynamics of futures on Deutsche mark, LIBOR and S&P500. En P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, & J. J. Dongarra (eds.), *Computational Science — iccs 2002. Lecture Notes in Computer Science* (pp. 1201-1208). Springer. https://doi.org/10.1007/3-540-47789-6_127
- Duffie, D., & Stein, J. C. (2015). Reforming LIBOR and other financial market benchmarks. *Journal of Economic Perspectives*, 29(2), 191-212. <https://doi.org/10.1257/jep.29.2.191>
- Fagerström, J., Bång, M., Wilhelms, D., & Chew, M. S. (2019). LiSep LSTM: A machine learning algorithm for early detection of septic shock. *Scientific Reports*, 9(1), 15132. <https://doi.org/10.1038/s41598-019-51219-4>
- Jamshidian, F. (1997). LIBOR and swap market models and measures. *Finance and Stochastics*, 1(4), 293-330. <https://doi.org/10.1007/s007800050026>
- Kumar, J., Goomer, R., & Singh, A. K. (2018). Long short-term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters. *Procedia Computer Science*, 125, 676-682. <https://doi.org/10.1016/j.procs.2017.12.087>
- Leal, F., Molina, C., & Zilberman, E. (2020). *Proyección de la inflación en Chile con métodos de machine learning*. Banco Central de Chile
- Li, D., Zhang, Y., Gong, D., Pan, L., & Zhao, Y. (2020). Gas data prediction based on LSTM neural network. *MS&E*, 750(1), 012175. <https://doi.org/10.1088/1757-899X/750/1/012175>
- Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv*, 1605.05101.
- Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33(6), 497-505. <https://doi.org/10.1016/j.omega.2004.07.024>
- Pastor-Cobo, M., & Torre-Barrio, P. *Elaboración de un algoritmo predictivo sobre la evolución del precio de las criptomonedas* (tesis de pregrado, Universidad Complutense de Madrid). Repositorio Universidad Complutense de Madrid.
- Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS One*, 15(1), e0227222. <https://doi.org/10.1371/journal.pone.0227222>
- Restrepo, L. F., & González, J. (2007). From Pearson to Spearman. *Revista Colombiana de Ciencias Pecuarias*, 20(2), 183-192.
- Robles, H. P. (2019). Uso de redes neuronales para el pronóstico de la inflación en México. *Almanaque Económico*, 1, 69-74.
- Rodríguez-Vargas, A. (2020). *Forecasting Costa Rican inflation with machine learning methods*. Banco Central de Costa Rica.

- Tan, X. (2019). LIBOR prediction using genetic algorithm and genetic algorithm integrated with recurrent neural network. *2019 Global Conference for Advancement in Technology (GCAT)*. Bangalore, India. <https://doi.org/10.1109/GCAT47503.2019.8978299>
- Villada, F., Muñoz, N., & García, E. (2012). Aplicación de las redes neuronales al pronóstico de precios en el mercado de valores. *Información Tecnológica*, 23(4), 11-20. <https://doi.org/10.4067/S0718-07642012000400003>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29-39).
- Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019). AT-LSTM: An attention-based LSTM model for financial time series prediction. *IOP Conference Series: Materials science and engineering*, 569, 052037. <https://doi.org/10.1088/1757-899X/569/5/052037>