

Proyecto ML.
Matemáticas para Machine Learning.

**Análisis y réplica del paper: Colombian inflation
forecast using Long Short-Term Memory approach.**

12 de julio de 2024

Juan Camilo Quintero.
Raul Torres.
Nicolás Naranjo.
Elizabeth Suárez Sánchez.

Índice general

1. Resumen (Abstract).	2
2. Introducción.	3
3. Descripción del Artículo Original.	4
4. Data.	5
5. Metodológico.	8
5.1. Long Short Term Memory	9
5.2. Ajuste de hiperparámetros.	12
5.3. Optimización Bayesiana.	13
5.4. Criterio de Optimización.	14
5.5. Dos modelos diferentes de aplicación del LSTM.	14
5.6. Modelo ARIMA.	16
6. Resultados.	17
7. Discusión y Conclusiones.	23
8. Referencias	24
Referencias	25

Capítulo 1

Resumen (Abstract).

En el artículo seleccionado para este proyecto, los autores buscaron predecir la inflación en Colombia utilizando redes neuronales Long Short-Term Memory (LSTM). Para ello, emplearon dos enfoques: en el primero, solo se utilizó información de la variable objetivo, la inflación; mientras que en el segundo, se incorporó información adicional proveniente de algunas variables relevantes.

Además, se utilizó el método de Rolling sample" dentro del proceso tradicional de construcción de las redes neuronales, seleccionando los hiperparámetros con criterios de minimización del error de pronóstico. Los resultados del artículo mostraron una mejor capacidad de pronóstico de la red bajo el segundo enfoque en comparación con el primero y con los modelos ARIMA optimizados para pronóstico (con o sin variables explicativas).

El artículo también demostró que la capacidad de pronóstico es más pronunciada en horizontes más largos, específicamente entre el séptimo y el duodécimo mes (Julian Alonso Cardenas-Cardenas, 2023).

Capítulo 2

Introducción.

La idea de este proyecto consiste, en primer lugar, en entender los conceptos teóricos abordados en el artículo y, principalmente, en comprender la aplicación de redes neuronales para resolver un problema. En este caso, se trata de un problema muy relevante para el Banco de la República: la predicción de la inflación.

Además, se intentará replicar las metodologías propuestas en el artículo, tanto para poner en práctica lo aprendido en el curso como para observar de forma tangible la aplicabilidad de las redes neuronales en diversas áreas.

La motivación del artículo surge de la importancia de la inflación en el comportamiento de los precios. Sin duda, la inflación es una variable crítica en cualquier economía, ya que influye en el poder adquisitivo de la población, en las decisiones de inversión de diferentes agentes y, en general, afecta tanto variables macroeconómicas como microeconómicas.

En Colombia, el Banco de la República (BR) es responsable de la política monetaria, lo que incluye acciones para mantener una inflación baja y estable, así como para lograr niveles sostenibles de producción y empleo bajo un esquema de metas de inflación. Por estas razones, es crucial para esta institución contar con proyecciones precisas de la inflación, ya que facilitan su correcta interpretación y la toma de decisiones adecuadas en materia de política monetaria.

Capítulo 3

Descripción del Artículo Original.

Este paper implementa redes neuronales de memoria a largo y corto plazo (LSTM), una técnica de Deep learning, para pronosticar la inflación general de Colombia con un año de anticipación mediante dos enfoques. El primero utiliza únicamente la información de la variable objetivo, mientras que el segundo incorpora información adicional de algunas variables relevantes. Empleamos el muestreo rodante en el proceso tradicional de construcción de redes neuronales, seleccionando los hiperparámetros con criterios para minimizar el error de pronóstico. Los resultados mostraron una mejor capacidad de pronóstico de la red con información de variables adicionales, superando tanto a la otra aplicación LSTM como a los modelos ARIMA optimizados para el pronóstico (con y sin variables explicativas). Esta mejora en la precisión del pronóstico es más pronunciada en horizontes de tiempo más largos, específicamente a partir del séptimo mes ([Julian Alonso Cardenas-Cardenas, 2023](#)).

Capítulo 4

Data.

En este artículo, los datos se centran en la inflación general en Colombia, medida a través de la variación anual del Índice de Precios al Consumidor (IPC), calculado mensualmente por el Departamento Administrativo Nacional de Estadística (DANE). Se utilizó una muestra mensual desde enero de 2002 hasta diciembre de 2019.

Para evitar el comportamiento atípico de la inflación causado por la pandemia de COVID-19, la muestra para la construcción del modelo LSTM (Long Short-Term Memory) termina antes del inicio de la pandemia. Los autores del artículo consideran que, una vez se superen los choques causados por la pandemia y otros eventos altamente relevantes, como el conflicto entre Rusia y Ucrania, la inflación volverá a un comportamiento más tradicional. Por lo tanto, se realizó un ejercicio separado para evaluar los modelos únicamente para el período de enero de 2020 a diciembre de 2022.

En una de las dos aplicaciones del modelo LSTM, se incluyeron variables adicionales que los autores consideraron relevantes para guiar el pronóstico a corto plazo (hasta un año). Estas variables son: la Tasa Representativa del Mercado (TRM) ([de la Republica, 2023](#)) del Peso Colombiano frente al Dólar Estadounidense, promedio mensual, calculada por el Banco de la República con información de la Superintendencia Financiera de Colombia; el Producto Interno Bruto Real de Colombia, ajustado estacionalmente y corregido por efectos de calendario (PIB), calculado por el DANE; y la brecha de precipitación en el país, calculada con datos de la Administración Nacional Oceánica y Atmosférica de los Estados Unidos (NOAA). Toda la información de estas variables se obtuvo de forma libre de la pagina del DANE ([DANE, 2023](#)) , del Banco de la República y la NOAA.

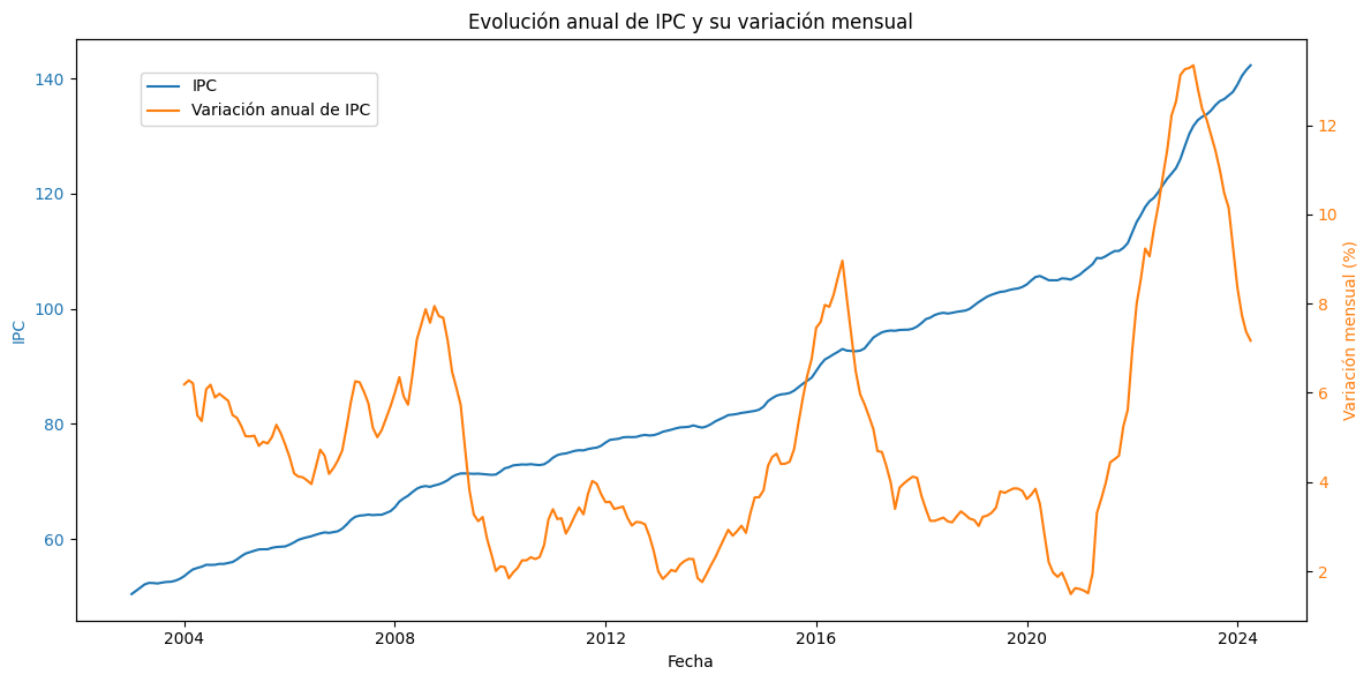


Figura 4.1: Índice de precios al consumidor.

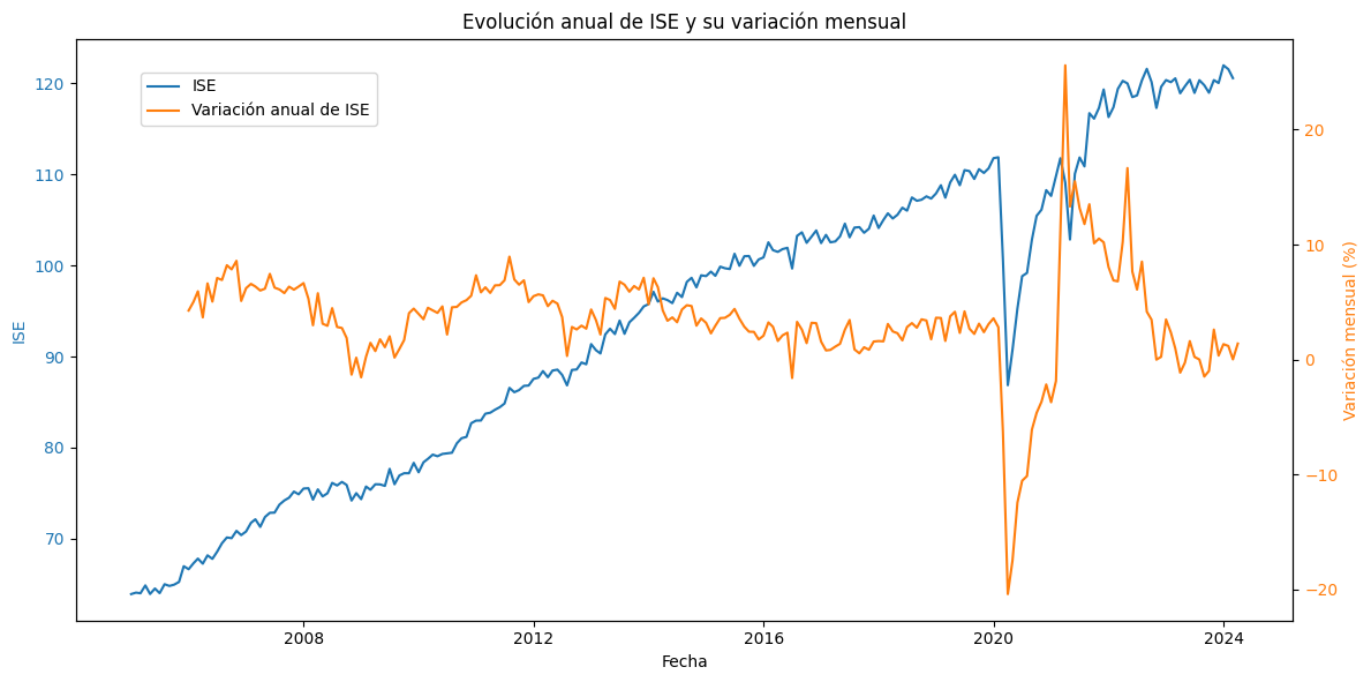


Figura 4.2: Indicador de seguimiento a la economía.

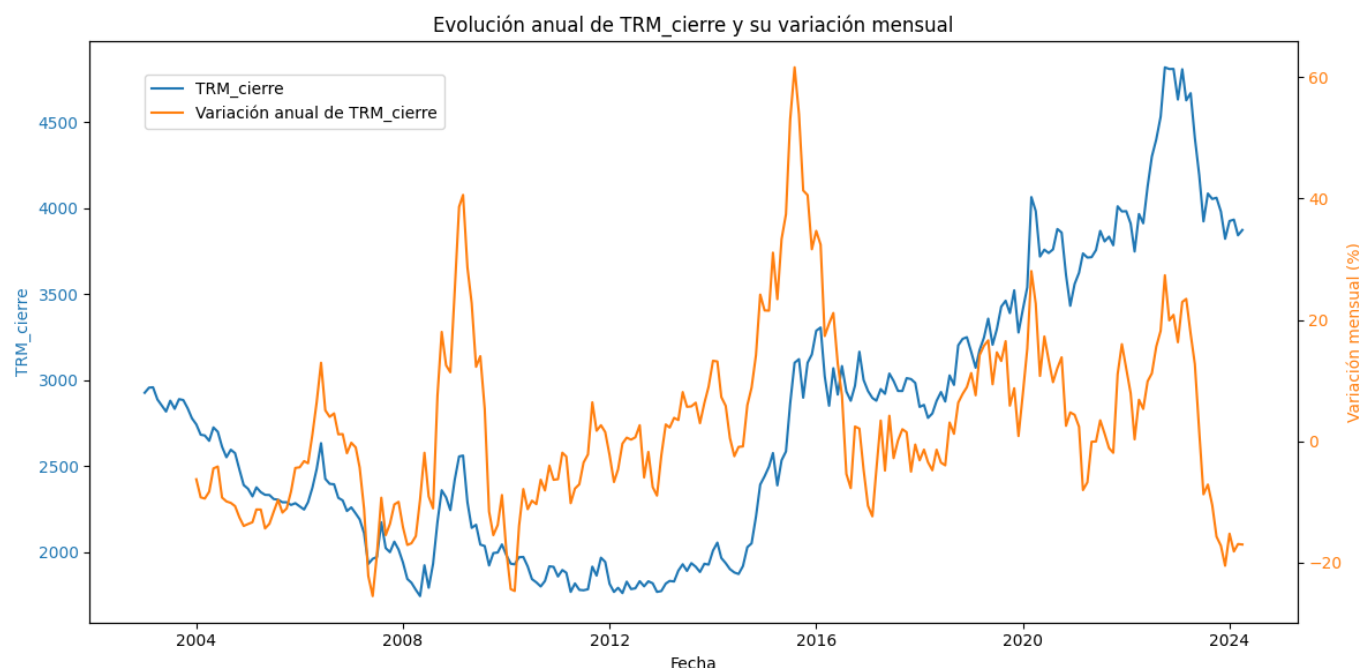


Figura 4.3: Tasa representativa del mercado.

La selección de estas tres variables se basa en su impacto relevante en el comportamiento de la inflación en Colombia. La TRM puede afectar el IPC especialmente debido a la transmisión de los precios de bienes e insumos importados (costos), al efecto de la competencia entre bienes nacionales y extranjeros (maximización de la utilidad) y porque actúa como un condensador de varios choques externos que también pueden afectar la inversión, la confianza del consumidor, las expectativas, etc.

El PIB, por otro lado, proporciona una imagen del estado de la demanda en la economía. Finalmente, la brecha de precipitación captura los choques de oferta más frecuentes y relevantes que afectan la inflación general en Colombia. Si bien existen otras variables que podrían ser importantes para predecir la inflación, en el artículo se decidió centrarse en estas tres variables por simplicidad y considerando que estas tres variables pueden, de alguna manera, reflejar los efectos de variables no consideradas.

Capítulo 5

Metodológico.

La metodología principal de este artículo consiste en la implementación de la red neuronal LSTM. Sin embargo, el enfoque es más complejo, ya que el modelo LSTM se aplica de dos formas diferentes: la primera utilizando únicamente los datos de inflación para hacer predicciones, y la segunda incorporando variables auxiliares junto con los datos de inflación. Además, se compara este modelo de aprendizaje automático con dos modelos tradicionales de pronóstico, ARIMA y ARIMAX. Para evaluar cuál es mejor, se utilizó la raíz del error cuadrático medio fuera de muestra (RMSE). Además, el modelo LSTM empleó la optimización bayesiana para encontrar el mejor ajuste de los hiperparámetros.

La siguiente gráfica explica mejor lo anterior:

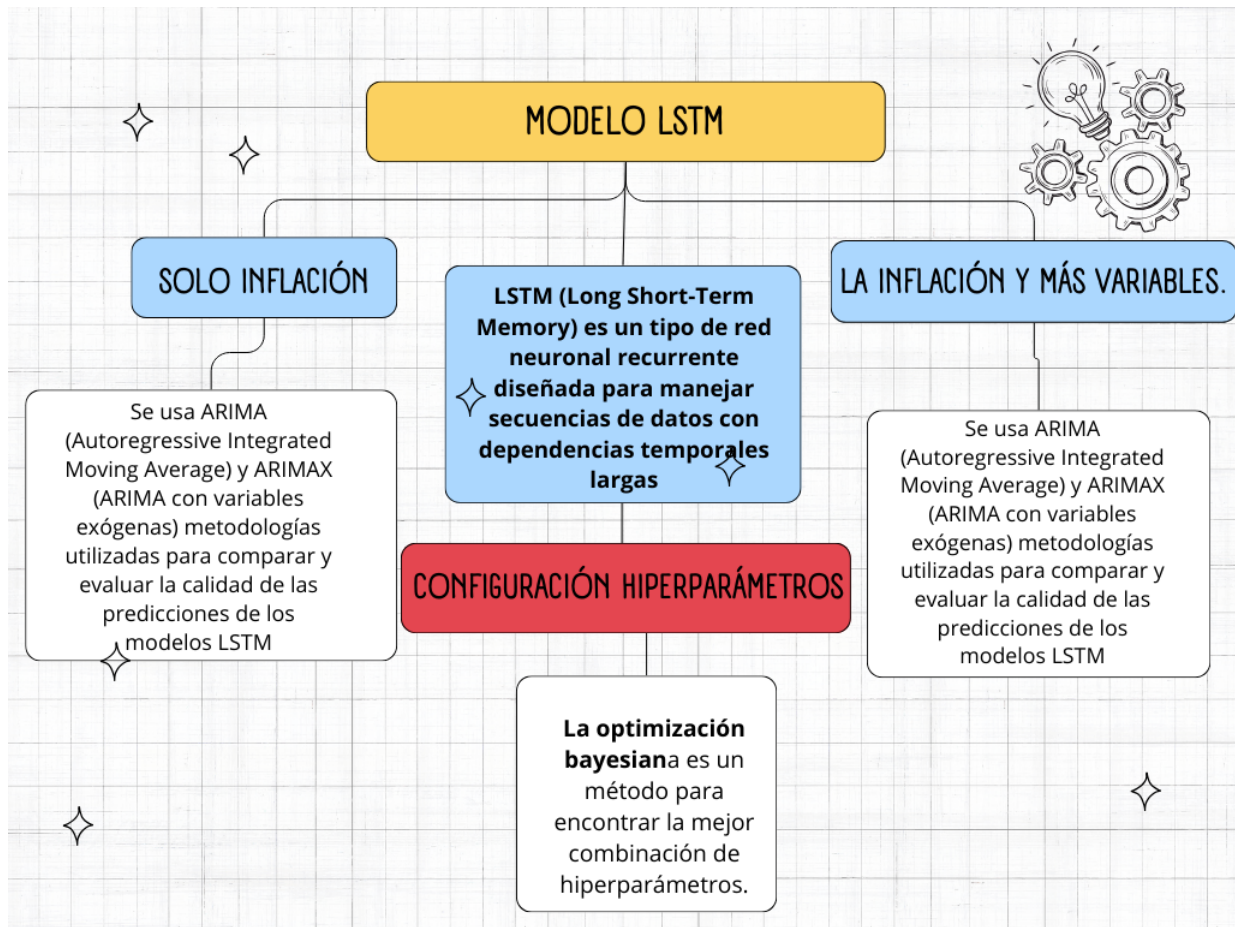


Figura 5.1: Mapa conceptual de la metodología.

5.1. Long Short Term Memory

Los modelos de redes neuronales artificiales han ganado mucha fuerza en los últimos años, impulsados por el desarrollo de la capacidad computacional y su enorme aplicabilidad en diversas áreas. Estos modelos buscan replicar el funcionamiento interno de una red neuronal biológica, simulando las conexiones que se forman para la toma de decisiones. En términos estadísticos, estos modelos extraen información relevante mediante la identificación de correlaciones e interacciones entre variables de manera no lineal, lo que los convierte en una clase particular de modelos parametrizados no lineales.

La estructura de una red neuronal artificial (Artificial Neural Network, ANN) generalmente está conformada por al menos tres capas de nodos a través de los cuales la información es procesada: nodos de entrada (input nodes o input layer), nodos ocultos (hidden nodes o

hidden layer) y nodos de salida (output nodes o output layer). La capa de entrada proporciona la información al modelo, las capas ocultas establecen las relaciones relevantes entre los datos para optimizar la respuesta en la capa de salida (como se muestra en la figura). El número de capas ocultas y los nodos en cada capa dependen de las características de los datos y sus relaciones.

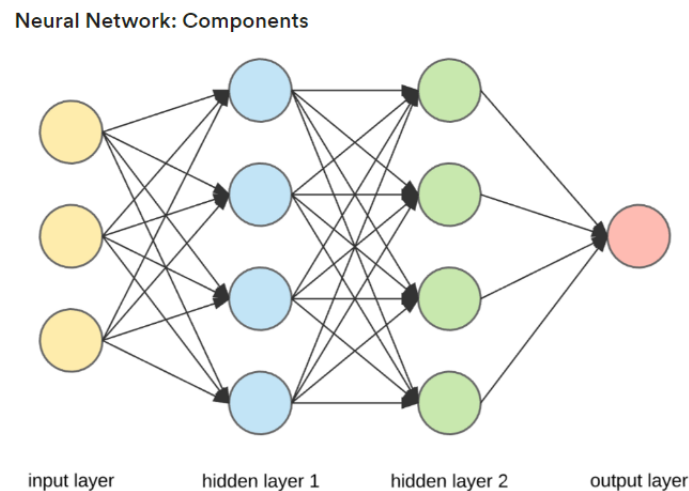


Figura 5.2: Imagen de una red neuronal (UP_GRAD, 2023)

Dentro de las ANN están las redes neuronales recurrentes RNN y dentro a su vez están las redes de memoria a largo y corto plazo (LSTM, por sus siglas en inglés) están diseñadas para aplicaciones en las que la entrada es una secuencia ordenada, donde el orden en que se ingresa la información es crucial. Las LSTM son un tipo de red neuronal recurrente que reutiliza las salidas como entradas. Esto significa que las salidas de la red neuronal se vuelven a ingresar como un nuevo input. Además, los nodos internos de la red no solo son recurrentes, sino que también almacenan y recuperan información, funcionando como una memoria. Finalmente, el valor del input, la salida anterior y los estados internos de almacenamiento se utilizan en los cálculos de los nodos.

A diferencia de las RNN normales, las LSTM tienen tres puertas diferentes que controlan el flujo de información (puertas de entrada, salida y olvido). Estas interactúan de una manera que incorpora parte de la información de la memoria a lo largo del tiempo y elimina la información no informativa para describir el comportamiento de la serie. De esta manera, las LSTM mantienen la estructura en cadena, pero el módulo repetitivo (célula) tiene una forma

diferente (Figura 3). La memorización del comportamiento de los datos es posible a través de las puertas y una línea de memoria. Esto nos permite mantener estados a largo plazo además de los estados a corto plazo.

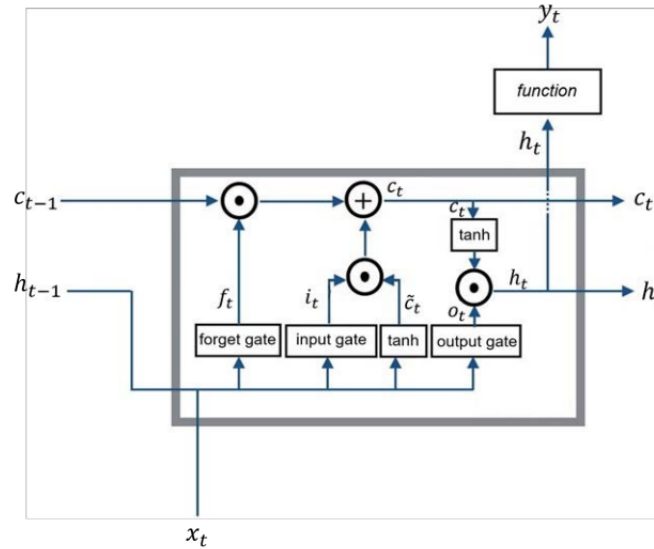


Figura 5.3: celda LSTM (Julian Alonso Cardenas-Cardenas, 2023)

Las ecuaciones que describen el proceso anterior y las planteadas por el artículo son:

$$\begin{aligned} f_t &= \sigma_g (W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma_g (W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma_g (W_o x_t + U_o h_{t-1} + b_o) \\ \tilde{C}_t &= \tanh (W_C x_t + U_C h_{t-1} + b_C) \end{aligned}$$

Figura 5.4: (Julian Alonso Cardenas-Cardenas, 2023)

donde x_t representa las variables de entrada (rezagos de la variable objetivo o de variables adicionales), W_f , W_i , W_o y W_C son las matrices de pesos que mapean la entrada de la capa oculta a las puertas y al estado de la célula de entrada, mientras que U_f , U_i , U_o y U_C son las matrices de pesos que conectan el estado de salida de la célula anterior con las tres puertas y el estado de la célula de entrada. Los b_f , b_i , b_o y b_C son cuatro vectores de sesgo. La σ_g es la función de activación, que normalmente es la función sigmoide, y la \tanh es la función tangente hiperbólica. El papel de estas funciones es filtrar parte de la información (útil/no útil) hacia el siguiente paso. Basado en los resultados de las cuatro ecuaciones anteriores, en cada iteración de tiempo t , el estado de salida de la célula, C_t , y la salida de la capa, h_t , se pueden calcular de la siguiente manera.

Figura 5.5: (Julian Alonso Cardenas-Cardenas, 2023)

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t$$

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

Figura 5.6: (Julian Alonso Cardenas-Cardenas, 2023)

5.2. Ajuste de hiperparámetros.

Bajo la estructura anterior, el siguiente paso es delinear el proceso por el cual se determinan los valores clave para el funcionamiento del modelo, estos valores son los hiperparámetros. Estos incluyen el número de capas ocultas, los nodos en cada capa, entre otros. Para lograr esto, se siguió un proceso iterativo (que incluye la optimización bayesiana) destinado a optimizar una función (en nuestro caso el RWMSE, raíz del error cuadrático medio ponderado), y en el cual otros valores afectan, como la tasa de aprendizaje, el número de épocas o el dropout (Prince, 2023).

En los algoritmos de aprendizaje automático y aprendizaje profundo, los hiperparámetros necesitan ser ajustados para lograr un rendimiento óptimo de los modelos. Los hiperparámetros se determinan encontrando sus valores óptimos a través de múltiples valores desconocidos mediante un procedimiento de optimización. En el artículo se enfocaron en ajustar el número de épocas, que se refiere a la cantidad de veces que todos los datos de entrenamiento pasan por la red para aprender de ellos; el número de nodos y capas ocultas (aplicación considera 1 y 2 capas ocultas debido a las características de la serie temporal a pronosticar); la tasa de aprendizaje, que controla la velocidad a la que el modelo se adapta al problema determinando la cantidad en que los pesos se actualizan durante el entrenamiento; y el dropout, que implica la probabilidad de eliminar ciertos nodos o neuronas de la red para verificar el rendimiento de una red más simple.

Los pasos que se realizaron en el artículo fueron los siguientes:

- i. Tomar la muestra (de entrenamiento) de enero de 2002 a diciembre de 2015 y un conjunto de posibles valores para los hiperparámetros.
- ii. Entrenar el modelo y calcular las predicciones doce pasos hacia adelante. Calcular y almacenar el RMSE por horizonte.
- iii. Agregar un mes a la muestra de entrenamiento y repetir el procedimiento hasta incluir diciembre de 2018 en la muestra (con pronósticos hasta diciembre de 2019).

- iv. Calcular el promedio del RMSE entre las muestras para cada horizonte de pronóstico. Calcular y almacenar el RWMSE asociado.
- v. Repetir el procedimiento cambiando los valores de los hiperparámetros hasta completar la cuadrícula definida en los párrafos siguientes.
- vi. Elegir el modelo con los valores de hiperparámetros que generaron el menor RWMSE.

En el artículo los autores expresan como buscaron encontrar un equilibrio entre un bajo RMSE y la eficiencia computacional, para lo cual plantearon un conjunto de hiperparámetros con dos capas ocultas (dado que la longitud de los datos no es muy grande) y para no comprometer la eficiencia computacional decidieron escoger un número de nodos entre 2 y 128. A partir de lo anterior evaluaron los siguientes valores : una y dos capas ocultas, con 4, 16, 32, 64 y 128 nodos en cada capa; para el número de épocas probamos con 20, 50, 100, 150 y 200; y para el porcentaje de dropout evaluamos con 0.4, 0.3 y 0.2. Tomaron la tasa de aprendizaje como fija (0.001) ([Julian Alonso Cardenas-Cardenas, 2023](#)).

Una vez obtenidos los resultados anteriores, se analizaron los patrones de combinaciones de hiperparámetros que generaron el menor RWMSE y se definió un nuevo rango en torno a los valores óptimos para cada hiperparámetro. Posteriormente, se volvió a ejecutar el algoritmo y se evaluó el RWMSE. Se encontró que, aunque la sensibilidad al ajuste de hiperparámetros variaba considerablemente en rangos amplios, esta disparidad disminuía significativamente al utilizar una cuadrícula más refinada alrededor de valores específicos. Esto permitió alcanzar un equilibrio entre precisión y eficiencia computacional.

5.3. Optimización Bayesiana.

Dentro de este artículo, los autores proponen utilizar la optimización bayesiana, los autores incluyeron este método porque evalúa exhaustivamente todas las combinaciones en algunos intervalos. Este procedimiento se centra en encontrar el mínimo de una determinada función $f(x)$ bajo un conjunto acotado, en este caso, el conjunto A. Los métodos bayesianos impulsan la búsqueda del conjunto óptimo de hiperparámetros al crear un modelo probabilístico donde el valor de la función objetivo función $f(x)$ es la métrica de validación del modelo, en nuestro caso, el RMSE ([Julian Alonso Cardenas-Cardenas, 2023](#)).

Por el momento no se abordará a profundidad este tema.

5.4. Criterio de Optimización.

La raíz del error cuadrático medio (RMSE) fuera de muestra es el criterio de rendimiento del modelo, este criterio se define según la siguiente ecuación, en donde N representa el tamaño de la muestra, y_t el valor observado en el tiempo t y \hat{y}_t la predicción de y_t .

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}$$

Figura 5.7: (Julian Alonso Cardenas-Cardenas, 2023)

También se usó una RMSE ponderada, dando más importancia a los horizontes más cortos. Con esta medida evaluaron el comportamiento y la predictibilidad, teniendo en cuenta el horizonte de pronóstico ($h = 1, \dots, 12$). Luego, la raíz del error cuadrático medio ponderado (RWMSE) para un modelo se define por la siguiente ecuación:

$$RWMSE_j = \frac{\sum_h \frac{1}{h} RMSE_{jh}}{\sum_h \frac{1}{h}}$$

Figura 5.8: (Julian Alonso Cardenas-Cardenas, 2023)

El RWMSE es el proceso de ajuste para seleccionar el mejor modelo para cada metodología, y el RMSE por horizonte de pronóstico para comparar entre los modelos seleccionados.

5.5. Dos modelos diferentes de aplicación del LSTM.

En este artículo se plantean los diferentes opciones de modelo:

- LSTM FM: El primer modelo que se propuso se basa únicamente en el historial de la variable a predecir (IPC). En este caso, las entradas de la red neuronal son los propios rezagos de la serie temporal, sin considerar el efecto de otras variables. Este modelo se puede encontrar replicado en el Github anexado con el nombre de la carpeta Proyecto-Final-ML

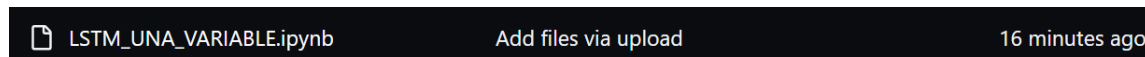


Figura 5.9

Dado que el código se encuentra bien comentado, a continuación se hace una explicación general de los pasos realizados en el código.:

- **Importar librerías y cambiar el directorio de trabajo** : En esta parte del código es donde se llaman todas las librerías a utilizar.
 - **Leer y preparar los datos** : En esta parte del código es donde se lee la información del excel Inflación extraído del Banco de la República.
 - **Escalar y dividir los datos** : En este paso se realizan dos cosas importantes, se reescalan los datos entre 0 y 1 y se divide la cantidad de datos entre los de entrenamiento (80 %) y los de prueba.
 - **Crear el conjunto de datos para el modelo** : En esta parte se organizan todas las variables para poder ser visualizadas gráficamente.
 - **Construir y entrenar el modelo LSTM** : Para la implementación del LSTM se usó Keras, una biblioteca de alto nivel para redes neuronales escrita en Python. Fue desarrollada con el objetivo de permitir una experimentación rápida y eficiente con redes neuronales profundas.
 - **Hacer predicciones y desescalar los datos**: En esta parte se hace la configuración para realizar las predicciones con los conjuntos de datos de entrenamiento y prueba.
 - **Calcular el error y graficar los resultados** : Se realizan dos procesos importantes calcular el error cuadrático medio de la raíz (RMSE) para las predicciones de entrenamiento y prueba, también se realizan las gráficas de los datos reales y las predicciones.
 - **Hacer predicciones futuras** : Se realiza la predicción para los siguientes 12 meses.
- • **LSTM EM**: En este segundo modelo, también se utilizó información de variables adicionales. La justificación de esta metodología es que el efecto de otras variables es relevante para el comportamiento futuro de la serie temporal de interés. El algoritmo funciona como se describió anteriormente, incluyendo también los rezagos de las variables explicativas como entradas.

5.6. Modelo ARIMA.

El modelo tradicional con el que se pone a prueba el LSTM es el modelo para la predicción de series temporales univariadas, es el modelo de promedio móvil autorregresivo integrado (ARIMA).

Por otro lado, Un modelo ARIMAX es el modelo ARIMA con un nuevo término para las variables exógenas (ecuación 7), en este caso se genera un vector que incluye las tres variables explicativas (presentadas anteriormente) con cierta cantidad de rezagos ([Julian Alonso Cardenas-Cardenas, 2023](#)).

No se profundizará en estos modelos, ya que son los modelos tradicionales que no tienen un componente aprendizaje de máquinas.

Capítulo 6

Resultados.

- **Resultados obtenidos en el artículo.**

La Figura 6.1 muestra los conjuntos de datos de entrenamiento y prueba, la línea negra representa el conjunto de entrenamiento y la línea roja el conjunto de prueba, ambos para el enfoque de ventana móvil (rolling approach). Los gráficos tienen una división basada en saltos de 7 meses para mostrar las diferencias en los datos de entrenamiento y prueba en diferentes períodos. Tomaron las últimas 12 observaciones como el conjunto de prueba, con la misma longitud que el número de horizontes a pronosticar ($h=12$). Dividieron los datos de esta manera debido a la relevancia del orden en los datos secuenciales.

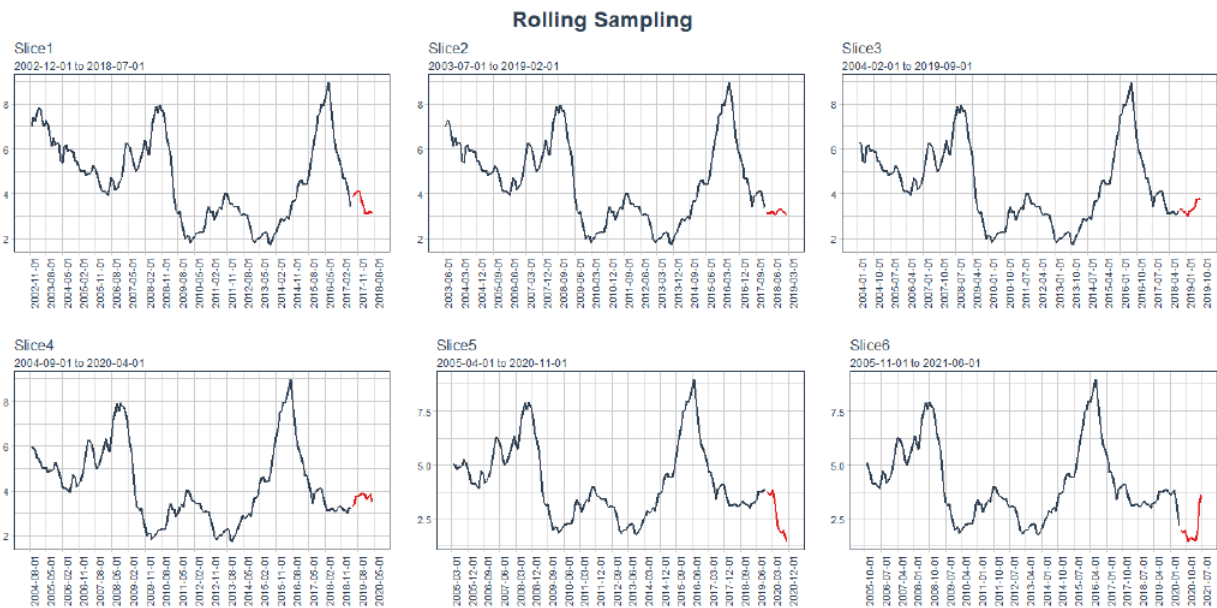


Figura 6.1: Ilustración del muestreo rodante. Notas: Cada diagrama (Slice) representa, como ejemplo, un paso del rolling sampling mensual. Cada paso se compone de una muestra de entrenamiento (línea negra) y su muestra de prueba de 12 meses (línea roja).

En el artículo también se analizó los factores que hacían mejor el rendimiento de la red neuronal. El número de nodos dentro de cada capa, el número de épocas (o iteraciones completas) y la tasa de dropout. La Figura 6.2 presenta el rendimiento del LSTM FM basado en un modelo con dos capas ocultas y diferentes números de épocas. Ambas capas ocultas incluyen diferentes especificaciones para el número de nodos. Los colores en la visualización indican la métrica RWMSE, donde las cajas más claras corresponden al RWMSE más bajo, y las más oscuras indican el RWMSE más alto. los resultados revelan que las predicciones más precisas se encuentran en el panel izquierdo (20 épocas), independientemente del número de nodos dentro de cada capa. En cuanto al número de nodos, el mejor rendimiento se obtiene con modelos que tienen pocos nodos en cada capa (cajas inferiores).

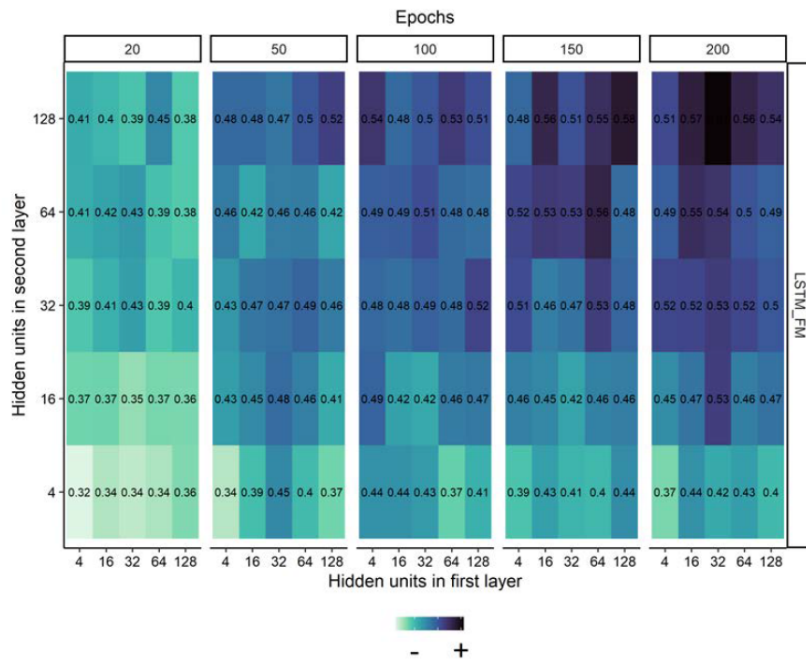


Figura 6.2: RWMSE para el LSTM FM con dos capas ocultas y diferentes números de nodos y épocas. Notas: Los números y los colores representan los valores de RWMSE. Los colores verde claro representan valores más bajos, mientras que los colores azul oscuro representan valores más altos.

Al enfatizar el número de capas ocultas, obtuvieron resultados similares o peores a medida que este número aumenta (6.3). El LSTM EM que usa dos capas ocultas requiere mucho poder computacional y, en general, sus resultados no son lo suficientemente buenos, mientras que para el LSTM FM no parece haber diferencias relevantes. Esta es la primera diferencia entre LSTM EM y LSTM FM; el rendimiento del primero se deteriora cuando se usa una arquitectura más compleja. En detalle, la Figura 6.3 muestra el RMSE para LSTM FM y LSTM EM en diferentes horizontes de pronóstico. El panel izquierdo muestra el RMSE para todos los horizontes usando una capa oculta, mientras que el panel derecho muestra el rendimiento con dos capas ocultas. En general, el modelo con una capa y variables explicativas es mejor en varios horizontes. Las diferencias más notables aparecen en horizontes más largos, donde el LSTM con variables explicativas y una capa oculta parece ser el mejor. Específicamente para horizontes largos, los resultados sugieren que las variables explicativas incluidas en la red proporcionan información valiosa para predecir la inflación en Colombia. En consecuencia, un modelo no tan complejo, con un número bajo de épocas y nodos, sería adecuado para pronosticar la serie temporal de inflación colombiana. Por esta razón, en el artículo una sola capa y siguieron con la selección del número de nodos y épocas (Figura 6.4) (Julian Alonso Cardenas-Cardenas, 2023).

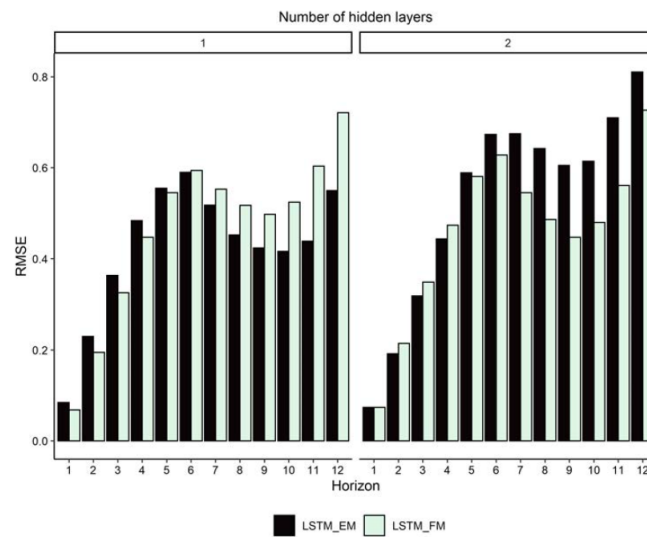


Figura 6.3: RMSE para LSTM FM y LSTM EM por horizonte de predicción.

La Figura 6.4 se construyó utilizando diferentes números de nodos para ambos modelos LSTM con una capa oculta. Demuestra nuevamente que un rango bajo de épocas es conveniente para estos datos, así como que solo se requiere un pequeño número de nodos para ambos modelos. Además, el número de épocas juega un papel importante solo en el LSTM FM; vemos que los modelos con muchos nodos empeoran cuando tienen más épocas. A diferencia del LSTM sin variables explicativas, el LSTM EM obtiene un rendimiento de pronóstico similar a lo largo del número de nodos y épocas (ver panel inferior del gráfico [\(Julian Alonso Cardenas-Cardenas, 2023\)](#)).

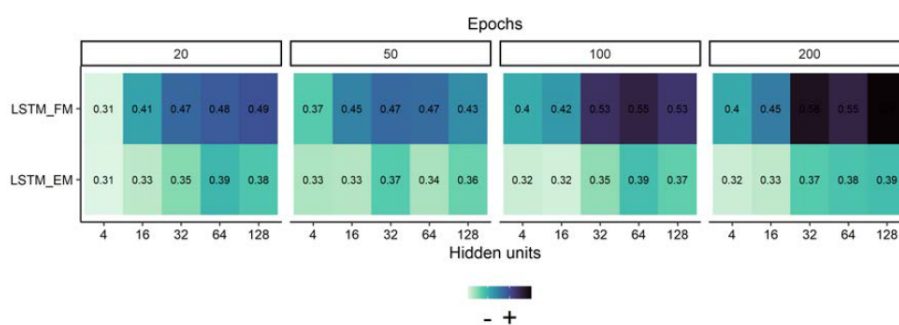


Figura 6.4: RWMSE para el LSTM FM y el LSTM EM con una capa oculta y diferentes números de nodos y épocas. Notas: Los números y los colores dentro del gráfico representan los valores de RWMSE. Los colores verde claro representan valores más bajos, mientras que los colores azul oscuro representan valores más altos.

Se evalúan los hiperparámetros de valores continuos alrededor de los valores encontra-

dos más plausibles según los resultados anteriores. Se ejecutó siguiendo la optimización bayesiana; la Tabla de la figura 6.6 que muestra la mejor combinación de valores de esos hiperparámetros con el mejor rendimiento según el RWMSE más bajo. Así, el algoritmo converge a valores cercanos a los encontrados previamente. Con estos valores, se ejecutaron los modelos LSTM y se comparó el RWMSE para cada modelo.

Hyperparameter	LSTM_FM	LSTM_EM
Hidden layers	1	1
Unit	13	6
Epoch	31	50
Dropout	0.2	0.1
Learning rate	0.001	0.001

Figura 6.5: Resultados de la Optimización Bayesiana.

Resultados obtenidos al replicar.

Una de las conclusiones importantes del artículo es que los LSTM funcionan mejor que los modelos ARIMA, particularmente para horizontes más largos.

La siguiente imagen muestra el RMSE para los dos modelos, LSTM considerando solo rezagos del IPC (LSTM FM), y el LSTM tomando información de variables explicativas (LSTM EM). Entre estos dos modelos, en general el LSTM EM presenta un mejor rendimiento. También compararon con los modelos ARIMA calculando el porcentaje de reducción en el RSME para cada modelo LSTM (ARIMA vs. LSTM FM, ARIMAX vs. LSTM EM, últimas dos columnas de la Tabla, respectivamente). Se observó que al usar LSTM-FM, hay una reducción en el RSME de alrededor del 7 % al 65 % en comparación con el modelo ARIMA, excepto para el primer horizonte. Además, la reducción en el RSME se vuelve progresivamente mayor a medida que el horizonte se alarga. Cuando se incluyen variables explicativas, el porcentaje de reducción del RMSE varía del 7 % al 68 %, mostrando una vez más una mayor reducción en horizontes más largos. Dado que el modelo de referencia es el modelo ARIMA en ambos casos, se pudo argumentar que para esta serie temporal, se obtuvo un mejor rendimiento en términos de bajo RMSE utilizando el método LSTM, especialmente en horizontes más largos ([Julian Alonso Cardenas-Cardenas, 2023](#)).

Horizon	(I)	(II)	(III)	(IV)	% Reduction	
	LSTM_FM	ARIMA	LSTM_EM	ARIMAX	(I) vs. (II)	(III) vs. (IV)
1	0.24	0.21	0.15	0.17	14.29	-11.76
2	0.41	0.44	0.25	0.31	-6.82	-19.35
3	0.53	0.70	0.36	0.41	-24.29	-12.20
4	0.63	0.95	0.44	0.50	-33.68	-12.00
5	0.72	1.24	0.57	0.61	-41.94	-6.56
6	0.77	1.49	0.60	0.68	-48.32	-11.76
7	0.77	1.66	0.58	0.75	-53.61	-22.67
8	0.76	1.86	0.53	0.83	-59.14	-36.14
9	0.76	2.04	0.46	0.92	-62.75	-50.00
10	0.76	2.20	0.39	0.98	-65.45	-60.20
11	0.80	2.29	0.35	1.02	-65.07	-65.69
12	0.83	2.36	0.34	1.05	-64.83	-67.62
RWMSE	0.50	0.87	0.32	0.45	-	-

Figura 6.6: RMSE por cada Horizonte de predicción

Capítulo 7

Discusión y Conclusiones.

Las siguientes conclusiones son respecto al artículo en estudio:

- Se observó que los modelos LSTM mejoran la capacidad de pronóstico de las especificaciones ARIMA. Por lo tanto, es razonable considerar la aplicación de estas técnicas novedosas en otras series económicas para problemas de pronóstico.
- Se encontró que aumentar la complejidad del modelo mediante la adición de capas ocultas o aumentando el número de nodos o épocas no necesariamente conduce a mejores resultados en términos de capacidad de pronóstico.
- Los resultados empíricos muestran que los modelos LSTM superan a los modelos ARIMA, siendo particularmente evidente en los horizontes de pronóstico más largos. Además, los resultados muestran la importancia de incluir variables auxiliares o explicativas potencialmente relacionadas con la serie a pronosticar, en este caso, la inflación.
- En términos del RMSE, se encontró una reducción del 7% para el segundo horizonte (mes de pronóstico) y del 65% para el horizonte 12 utilizando el LSTM con sus propios rezagos en comparación con el modelo ARIMA. Por su parte, el LSTM con variables explicativas muestra reducciones de hasta el 68% en el horizonte más largo en comparación con el modelo ARIMAX (y del 59% en comparación con el primer LSTM).

Capítulo 8

Referencias

Referencias

- DANE. (2023). *Bib*. Descargado de <https://www.dane.gov.co/>
- de la Republica, B. (2023). *Ipc*. Descargado de <https://www.banrep.gov.co/es/estadisticas/precios-e-inflacion>
- Julian Alonso Cardenas-Cardenas, e. J. C.-B. N. M.-C. (2023). *A state space representation of the chain ladder linear model*. Descargado de <https://repositorio.banrep.gov.co/server/api/core/bitstreams/4b785d6c-4bdd-4e09-b289-96d47a44a57c/content>
- Prince, S. J. (2023). *Understanding deep learning*. Descargado de <https://udlbook.github.io/udlbook/>
- UP_GRAD. (2023). *Neural network: Architecture, components top algorithms*. Descargado de <https://www.upgrad.com/blog/neural-network-architecture-components-algorithms/>