Recibido: 13/03/2022 / Aceptado: 20/07/2022
doi: https://doi.org/10.26439/interfases2022.n015.5817

APLICACIÓN DE MÉTODOS DE *DEEP LEARNING* EN SERIES DE TIEMPO PARA EL PRONÓSTICO DE LA SITUACIÓN MACROECONÓMICA EN AMÉRICA LATINA

Víctor Augusto Alegre Ibáñez
victoralegre@uni.pe / ORCID: 0000-0002-1456-8065
Universidad Nacional de Ingeniería (UNI), Lima, Perú

Jose Martin Lozano Aparicio
jlozano@uni.edu.pe / ORCID: 0000-0003-1086-5799
Universidad Nacional de Ingeniería (UNI), Lima, Perú

Resumen

Los métodos de *deep learning* pueden ser aplicados para generar modelos de pronóstico. Nosotros trabajamos con el producto bruto interno (PBI) de seis países de América Latina: Argentina, Brasil, Chile, Colombia, México y Perú empleando indicadores macroeconómicos anuales y trimestrales, del Banco Mundial y la Comisión Económica para América Latina y el Caribe (CEPAL), respectivamente. Para el preprocesamiento de los datos, a las series trimestrales se agregaron como características adicionales la descomposición de estas en tendencia, estacionalidad y residuo, con la finalidad de aportar más información a los modelos. Además, se reemplazaron datos atípicos producto del impacto de la pandemia del COVID-19 en la economía mundial. Se construyeron modelos de Perceptrón Multi Capa, Red Neuronal Convolucional, LSTM, GRU y SeqToSeq para cada país y frecuencia de sus series, y luego se evaluaron mediante validación cruzada continua y métricas MAE, RMSE y MAPE. Los modelos óptimos varían por cada caso

PALABRAS CLAVE: aprendizaje profundo / pronóstico de PBI / CEPAL / redes neuronales

APPLICATION OF DEEP LEARNING METHODS IN TIME SERIES FOR THE FORECAST OF THE MACROECONOMIC SITUATION IN LATIN AMERICA

Abstract

Deep learning methods can be applied to generate predictive models. We worked with the gross domestic product (GDP) of six Latin American countries: Argentina, Brazil, Chile, Colombia, Mexico, and Peru, using annual and quarterly macroeconomic indicators from the World Bank and the Economic Commission for Latin America and the Caribbean (ECLAC), respectively. For the pre-processing of the data, we decomposed the quarterly series into trend, seasonality, and residual and used them as additional characteristics to provide more information to the models. In addition, outliers resulting from the impact of the COVID-19 pandemic on the world economy were replaced. Multilayer perceptron, convolutional neural networks, LSTM, GRU, and SeqToSeq models were built for each country and their series' frequency, then evaluated by continuous cross-validation and MAE, RMSE, and MAPE metrics. The optimal models vary for each case.

KEYWORDS: deep learning / GDP forecasting / CEPAL / neural network

1. INTRODUCCIÓN

El *deep learning* (DL), parte del *machine learning* (ML), se caracteriza por sus arquitecturas de redes neuronales de varias capas ocultas a lo que se denomina profundidad (de ahí el nombre). Este tiene aplicaciones como el reconocimiento de patrones, el procesamiento de lenguaje natural y la predicción, de la cual trataremos en este trabajo.

La predicción puede ser muy útil para la toma de decisiones sobre eventos a futuro. Esto es respaldado por datos de sucesos anteriores como los datos históricos o series de tiempo. Ejemplos de esto son los pronósticos de precios de artículos, *commodities*, acciones, o los valores de indicadores económicos. Estos son importantes para una empresa o para un Gobierno, para vigilar y tomar medidas para mejorar su situación.

Los indicadores macroeconómicos nos describen la situación económica de un país o región, como también el bienestar de su población. El indicador que se pronosticará es el producto bruto interno (PBI) el cual mide el ingreso de todos los miembros de la economía, que es igual al gasto total en la producción de bienes y servicios finales. Esto aplica para un país en cierto periodo (Mankiw, 2014).

Para el pronóstico de series temporales, ya existen modelos estadísticos como las medias móviles y las autorregresiones. Pero en este trabajo nos enfocaremos en modelos de *deep learning* que también tienen buen desempeño con las series de tiempo, y los aplicaremos para predecir indicadores macroeconómicos de países de América Latina. Esta investigación comienza describiendo los conceptos principales en la sección 2. Los trabajos relacionados en la sección 3. La metodología usada es presentada en la sección 4. Los resultados y las discusiones son mostradas en la sección 5. Finalmente, presentamos nuestras conclusiones.

2. MARCO TEÓRICO

Las *series de tiempo* son datos secuenciales cuyos registros tienen un tiempo asociado (hora, fecha, año, entre otros). Por ejemplo, transacciones de compras. Una serie de tiempo es un tipo de dato secuencial en la que sus registros consisten en datos numéricos medidos sobre el tiempo (Tan et al., 2006). Estos pueden ser generados por procesos naturales y económicos como los mercados de acciones, observaciones científicas, médicas o fenómenos naturales (Han, 2012).

El *aprendizaje profundo* (*deep learning*) es un subgrupo del aprendizaje automático (*machine learning*) que pone énfasis en el aprendizaje de capas sucesivas de representaciones cada vez más significativas. El número de capas que contribuyen al modelo se llama profundidad, y estos modelos mayormente son redes neuronales (Chollet, 2018). Las redes neuronales de aprendizaje profundo usadas en este trabajo son:

- *Perceptrón multicapa (MLP)*, esta red neuronal más básica es usada para regresión lineal, en la cual los nodos de cada capa están conectados a los nodos de la siguiente capa (Skansi, 2018).
- *Redes neuronales convolucionales (CNN)* tienen una o más capas convolucionales que realizan la operación de convolución. Esta consiste en aplicar kernels o filtros por medio de producto escalar, para obtener mapas de características (Skansi, 2018).
- *Redes neuronales recurrentes (RNN)*, son una clase de redes neuronales donde las conexiones entre unidades forman un ciclo dirigido. Este crea un estado interno de la red, lo cual permite exponer un comportamiento temporal dinámico. Las RNN aprovechan su memoria interna para manejar secuencias de entradas, en vez de solo mapear entradas y salidas, son capaces de aprovechar una función de mapeo para las entradas sobre el tiempo, hacia una salida (Lazzeri, 2021).
- *Long short-term memory (LSTM)* modifica las RNN para resolver el problema del desvanecimiento de gradiente, haciéndolo capaz de aprender a largo plazo. La LSTM tiene mecanismos internos llamados puertas y estados de célula que pueden regular el flujo de información. La célula toma decisiones sobre qué, cuánto y cuándo guardar y liberar información. Estos aprenden cuándo permitir a la información entrar, salir o eliminar, a lo largo del proceso iterativo de hacer conjeturas, retropropagación de error y ajuste de pesos vía gradiente descendente (Lazzeri, 2020). Una variante de esta red neuronal es *gate recurrent unit* (GRU).
- *Seq2Seq (sequence-to-sequence)* es aquel que toma como entrada a una secuencia de objetos (ya sea de palabras, letras, series de tiempo, etc.) y da otra secuencia de objetos como salida. El modelo se compone de un codificador y decodificador. Codificador consiste en una RNN (también puede ser LSTM o GRU para evitar el desvanecimiento de gradiente). Esta toma la secuencia de entrada hasta generar el último estado oculto. En el caso de un codificador LSTM, también se considera el último estado de la célula. Decodificador consiste en una RNN, el último estado oculto del decodificador se usa como entrada de todas las células, y si fuera LSTM, el último estado de célula se usa en la primera célula del decodificador (Sutskever et al., 2014).

3. TRABAJOS RELACIONADOS

Cook y Smalter Hall (2017) predicen la tasa de desempleo la cual fue recolectada mensualmente por la Agencia Estadounidense de Trabajo y Estadísticas (US Bureau of

Labor and Statistics). Se usaron modelos basados en diferentes arquitecturas de redes neuronales (perceptrón multicapa [MLP], red neuronal convolucional [CNN], *long short-term memory* [LSTM] y *encoder-decoder* [con LSTM]), y se compararon con el modelo estadístico DARM. La métrica empleada fue el error medio absoluto (MAE). Se resalta que la arquitectura *encoder-decoder* supera a los modelos de comparación en todos los horizontes de predicción (hasta cuatro trimestres).

Jung et al. (2018) predicen el crecimiento del PBI real a corto plazo para 7 países de diferentes geografías y desarrollo económico (Alemania, México, Filipinas, España, Reino Unido. Estados Unidos y Vietnam). La fuente fue la base de datos World Economic Outlook (WEO) del Fondo Monetario Internacional. También se agregaron datos de índices de mercados de acciones, precios de energías, entre otros, proveídos por Bloomberg, y datos procedentes de la Guía Internacional de Riesgo País (ICRG). Se emplearon técnicas de *machine learning* como *elastic nets*, *super learner* y RNN. La meta del artículo fue evaluar si estas técnicas podían mejorar la precisión en la predicción, concluyendo que superan a los empleados por la WEO.

Viswanath et al. (2019) proponen pronosticar los monzones en la región central de la India por medio de clasificación, usando métodos de *deep learning* como LSTM y Seq2Seq, y comparando con métodos tradicionales como SVM y KNN. Los datos fueron mediciones espaciotemporales de la región central de la India, la cual considera las lluvias diarias durante junio a septiembre, desde los años de 1948 al 2014, cuya fuente fue el Departamento Meteorológico de la India. El trabajo concluye que los modelos propuestos se desempeñan mejor que los modelos de clasificación SVM y KNN.

Nguyen y Nguyen (2020) proponen un modelo de pronóstico del PBI por medio de *transfer learning*. Para este propósito se implementaron modelos basados en LSTM y *encoder-decoder*, con y sin *transfer learning*. Los datos se obtuvieron de JSTdatasetR4 y están comprendidos de 17 variables macroeconómicas de 17 países desarrollados, lo cual dividen en 2 grupos: datos de 16 países para el entrenamiento, y el restante para afinamiento. Se evaluaron los modelos con las métricas: RMSE, nRSME, MAPE, sMAPE; que, según los resultados, los modelos con *transfer learning* tienen mejor desempeño en el preentrenamiento y en el afinamiento.

Kelany et al. (2020) emplean modelos basados en LSTM para pronosticar precios futuros para acciones de bajo, mediano y alto riesgo, y los comparan con regresión logística y *random forest*. Se usaron las métricas MAE y RMSE para las evaluaciones, de las cuales el modelo LSTM supera a los otros dos modelos.

Zyatkov y Krivorotko (2021) construyen un indicador para pronosticar el comienzo de recesiones en la economía de EE. UU. usando métodos de *machine learning*. Para este propósito se emplearon variables socioeconómicas las cuales son típicas en

periodos precrisis. Los métodos empleados fueron KNN, SVM, *random forest*, perceptrón multicapa y LSTM. Los horizontes para el indicador son los siguientes 6, 12 y 24 meses.

4. METODOLOGÍA

En esta sección, se explican los procesos llevados a cabo para construir y entrenar los modelos de predicción.

4.1 Recolección de datos

Se recolectaron series de tiempo de indicadores macroeconómicos de 6 países Latinoamericanos: Argentina, Brasil, Chile, Colombia, México y Perú.

- Anuales. Consisten en 14 series y 360 observaciones (en total 5040 datos), las cuales son el PBI real, componentes de consumo del PBI, valores agregados de industria y agricultura, precios de *commodities*. Están medidos en dólares americanos (USD) y a precios constantes de 2010, para que los indicadores no estén afectados por la inflación. Las observaciones empiezan desde el año 1960 hasta 2020. Las series se obtuvieron de la base de datos del Banco Mundial (2021). La descripción de las características se muestra en la tabla 1.
- Trimestrales. Son 11 series y 512 observaciones (en total 5632 datos), conformadas por el PBI real y por sectores. Medidos en volumen de moneda nacional y por año base para que no sean afectados por la inflación. El año base y periodo de observación depende de cada país: Argentina año base 2004 y periodo del 2004-2021, Brasil año base 2010 y periodo de 1996-2021, Chile año base 2013 y periodo 1996-2021, Colombia año base 2015 y periodo 2005-2021, México año base 2013 y periodo 1993-2021, y Perú año base 2007 y periodo 2007-2021. Los datos fueron recolectados de la web de la Comisión Económica para América Latina y el Caribe (2021). Las características se describen en la tabla 2.

Tabla 1

Características de las series anuales

Característica	Descripción
GDP	Producto bruto interno real (en USD, 2010).
CONS_TOTAL	Gasto de consumo final (consumo privado más Gobierno) (en USD, 2010).
INVEST	Formación bruta de capital (inversión) (en USD, 2010).
EXPORT	Exportaciones de bienes y servicios (en USD, 2010).
IMPORT	Importaciones de bienes y servicios (en USD, 2010).
AGRICUL	Agricultura, valor agregado (en USD, 2010).

(*continúa*)

(*continuación*)

INDUST	Industria, valor agregado (en USD, 2010).
INFLA_GDP_DEF	Inflación, deflactor del PBI (en porcentaje).
ENERGY	Índice anual real de energía (carbón, crudo de petróleo y gas natural) (en USD, 2010).
AGRICULTURE	Índice anual real de agricultura (alimentos, bebidas, materias primas agrícolas).
FERTILIZERS	Índice anual real de fertilizantes (en USD, 2010).
METMIN	Índice anual real de metales y minerales (aluminio, cobre, hierro, plomo, níquel, zinc) (en USD, 2010).
PRECIOUSMET	Índice anual real de metales preciosos (oro, plata y platino) (en USD, 2010).
HMUV	Índice deflactor de precios de *commodities* (en porcentaje, 2010: 100 %).

Nota. Series obtenidas del Banco de Datos del Banco Mundial, 2021 (Recuperado el 24 de octubre del 2021, de https://databank.bancomundial.org/).

Tabla 2

Características de las series trimestrales

Característica	Descripción
GDP	Producto bruto interno (en moneda local).
AGRO	Agricultura, ganadería, caza, silvicultura y pesca (en moneda local).
MINERIA	Explotación de minas y canteras (en moneda local).
INDUS	Industrias manufactureras (en moneda local).
SUMELECAGUA	Suministro de electricidad, gas y agua.
CONSTR	Construcción.
COMER	Comercio al por mayor y al por menor, reparación de bienes, hoteles y restaurantes.
TRANSPCOM	Transporte, almacenamiento y comunicaciones.
INMOBIL	Intermediación financiera, actividades inmobiliarias, empresariales y de alquiler.
GOBIERNO	Administración pública, defensa, seguridad social obligatoria, enseñanza, servicios sociales.
IMPUESTOS	Impuestos.

Nota. Series obtenidas de CEPALSTAT: Bases de Datos y Publicaciones Estadísticas, por la Comisión Económica para América Latina y el Caribe, 2021 (Recuperado el 14 de diciembre del 2021 de https://statistics.cepal.org/portal/cepalstat/dashboard.html).

4.2 Preprocesamiento

El siguiente preprocesamiento de los datos nos permitirá tener mejores resultados, el cual se divide en cuatro fases:

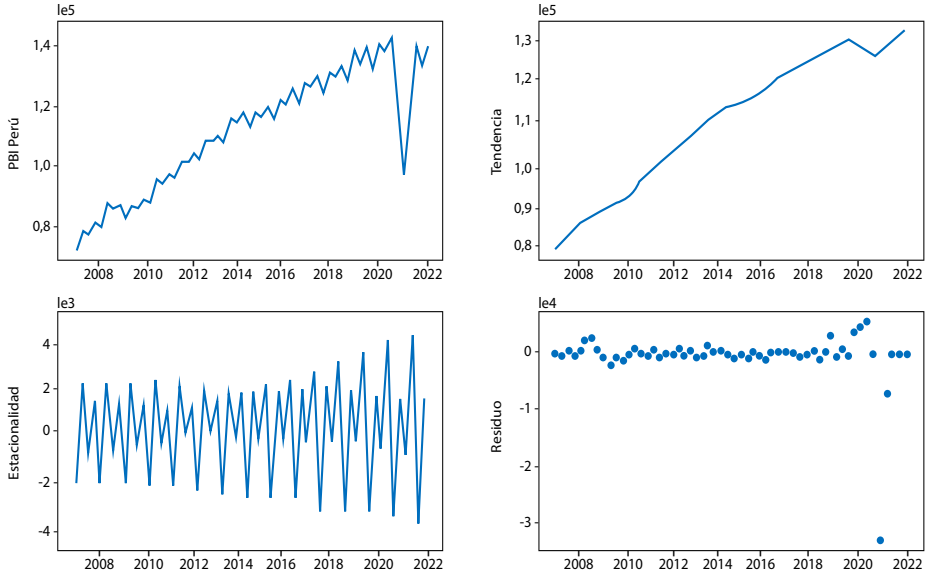
4.2.1 Descomposición de series temporales

Se descomponen las series trimestrales (y_t) en sus tres componentes: tendencia (T_t), estacionalidad (S_t) y residuo (r_t) (ecuación 1). Estas se adicionarán al conjunto de datos como características. Para este propósito se aplica la técnica de descomposición STL (*seasonal and trend decomposition using loess*) debido a que es robusta para la detección de datos atípicos (Cleveland et al., 2021). En la figura 1 se ejemplifica este proceso.

$$y_t = T_t + S_t + r_t \tag{1}$$

Figura 1

Ejemplo de descomposición de serie trimestral



Nota. Resultado de descomponer la serie del PBI trimestral de Perú aplicando la técnica de descomposición STL.

4.2.2 Reemplazo de datos atípicos

Se detectan los datos atípicos de los residuos aplicando la regla del rango inter-cuartil (IQR^r y luego reemplazándolos por cero, como lo vemos en las ecuaciones 2 a 5. Con los residuos resultantes (r'), se recomponen las series (y'_t) según la ecuación 6. Esta nueva serie reemplazará a la original en el conjunto de datos. Con este proceso retiramos el impacto negativo de la pandemia del COVID-19. Un ejemplo de este proceso lo observamos en la figura 2.

$$IQR^r = Q_3^r + Q_1^r \tag{2}$$

$$l = Q_1^r - 1.5 \times IQR^r \tag{3}$$

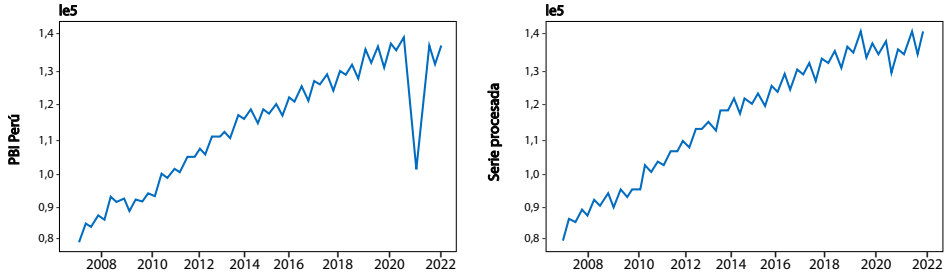
$$u = Q_3^r + 1.5 \times IQR^r \tag{4}$$

$$r' = f(r) = \begin{cases} 0, & (r < l) \vee (r > u) \\ r, & l \leq r \leq u \end{cases} \tag{5}$$

$$y_t' = T_t + S_t + r_t' \tag{6}$$

Figura 2

Ejemplo de reemplazo de datos atípicos en series trimestrales



Nota. Resultado de aplicar la técnica a la serie del PBI trimestral de Perú.

4.2.3 Escalamiento mínimo-máximo

Los datos de entrada y salida se escalan por mínimo y máximo hacia el rango (0, 1) para el correcto funcionamiento de los algoritmos de aprendizaje. La ecuación 7 muestra cómo se calcula.

$$x' = \frac{x - x_{mín.}}{x_{máx.} - x_{mín.}} \tag{7}$$

4.2.4 Generación de características pasadas y futuras

Se generan características de observaciones de *L* pasos de tiempo anteriores *(lag features*) a partir de los datos de entrada *X* que se usarán para pronosticar *H* pasos a futuro (*future features*) de los datos de salida *Y*. Para el estudio, se pronosticarán solo

H = 1 paso de tiempo a futuro y el dato de salida *Y* será la serie del PBI, esto será de un año para las series anuales, y de un trimestre para el otro caso. El número de pasos a pasado *L* se encontrarán por ajuste de hiperparámetros. La ecuación 8 muestra cómo serán los datos de entrada y salida.

$$(X_t, Y_t) \rightarrow (X_{t-L}, \cdots, X_{t-1}, X_t, Y_{t+H}) \tag{8}$$

4.3 Modelos e hiperparámetros

Para la construcción de los modelos anuales se tomaron 13 características de entrada y una de salida (el PBI a pronosticar), para los trimestrales son 43 de entrada y 1 de salida. Se fijaron los siguientes hiperparámetros para todos los modelos: la función de activación ReLU, el optimizador Adam, la función de pérdida que corresponde al error cuadrático medio (MSE) y la condición de parada de 50 épocas después de encontrar el menor error de validación. Estos parámetros fueron considerados según lo aplicado en los trabajos de Zyatkov y Krivorotko (2021) y Viswanath et al. (2019).

Respecto a los hiperparámetros con espacios de búsqueda presentes en las tablas 3 a 7 se detalla lo siguiente: se consideraron los números de pasos pasados en un rango de 4 a 20 para encontrar el óptimo de forma experimental; de la misma forma se consideraron los números para capas ocultas. Para el número de nodos, kernels y células se eligieron múltiplos de 32, y la ratio de aprendizaje son números pequeños (potencias entre 10^{-2} y 10^{-4}), basándose en el trabajo de Guimarães (2022).

Para la búsqueda de los hiperparámetros óptimos se empleó la técnica de hiperbanda, propuesto por Li et al. (2018), donde recursos predefinidos como iteraciones, muestras o características son asignados a muestras aleatorias de configuraciones de hiperparámetros, de los cuales se buscan los óptimos por medio de *successive halving*, técnica propuesta por Jamieson y Talwalkar (2016), la cual descarta las configuraciones con peores métricas y asigna más recursos a las mejores por medio de iteraciones. A continuación, se detalla la construcción de los modelos y las tablas 4 a 7, las cuales indican el espacio de búsqueda y los valores óptimos encontrados para las series anuales y trimestrales.

4.3.1 Perceptrón multicapa (MLP)

En la tabla 3 se muestran los hiperparámetros a buscar, como el número de pasos pasados necesarios para pronosticar, el número de capas ocultas, el número de nodos por capa oculta y la ratio de aprendizaje.

Tabla 3

Hiperparámetros para modelos basados en MLP

Hiperparámetro	Espacio de búsqueda	Óptimo anual	Óptimo trimestral
Número de pasos pasados	[4, 6, 8, 10, 12, 14, 16, 18, 20]	6	4
Número de capas ocultas	[1, 2, 3, 4, 5]	5	3
Número de nodos por capa	[32, 64, 96, 128, 160]	64	96
Ratio de aprendizaje	[10^{-2}, 10^{-3}, 10^{-4}]	10^{-2}	10^{-2}

4.3.2 Red neuronal convolucional (CNN)

La tabla 4 indica que se buscan el número de pasos pasados óptimo para este modelo, el número de kernels para la capa convolucional (la cual será una sola capa y unidimensional), el tamaño de kernel y *max pooling*, como también el número de nodos de capa densa, que procesarán la salida de la capa convolucional. También se busca la ratio de aprendizaje.

Tabla 4

Hiperparámetros para modelos basados en CNN

Hiperparámetro	Espacio de búsqueda	Óptimo anual	Óptimo trimestral
Número de pasos pasados	[4, 6, 8, 10, 12, 14, 16, 18, 20]	4	4
Número kernels capa convolucional	[32, 64, 96, 128, 160]	128	128
Tamaño de kernel	[2, 3]	2	2
Tamaño de *max-pooling*	_	2	2
Número de nodos capa densa	[32, 64, 96, 128, 160]	128	96
Ratio de aprendizaje	[10^{-2}, 10^{-3}, 10^{-4}]	10^{-2}	10^{-3}

4.3.3 *Long short-term memory* (LSTM)

La tabla 5 indica que se buscarán el número de pasos pasados, el número de células por capa, la ratio de aprendizaje. Se usarán dos capas LSTM.

Tabla 5

Hiperparámetros para modelos basados en LSTM

Hiperparámetro	Espacio de búsqueda	Óptimo anual	Óptimo trimestral
Número de pasos pasados	[4, 6, 8, 10, 12, 14, 16, 18, 20]	4	4
Número de capas LSTM	_	2	2
Número de células por capa	[32, 64, 96, 128, 160]	128	160
Ratio de aprendizaje	[10^{-2}, 10^{-3}, 10^{-4}]	10^{-2}	10^{-3}

4.3.4. Gated recurrent unit (GRU)

En la tabla 6 se muestrans los hiperparámetros a buscar: número de pasos pasados, número de células por capa y ratio de aprendizaje. Se fija el número de capas GRU a dos.

Tabla 6

Hiperparámetros para modelos basados en GRU

Hiperparámetro	Espacio de búsqueda	Óptimo anual	Óptimo trimestral
Número de pasos pasados	[4, 6, 8, 10, 12, 14, 16, 18, 20]	6	4
Número de capas GRU	_	2	2
Número de células por capa	[32, 64, 96, 128, 160]	32	160
Ratio de aprendizaje	$[10^{-2}, 10^{-3}, 10^{-4}]$	10^{-2}	10^{-3}

4.3.5 *Sequence to sequence* (Seq2Seq)

La tabla 7 indica el número de pasos pasados a buscar, como también el número de células por capa, ya sea para el codificador y decodificador, y por último la ratio de aprendizaje.

Tabla 7

Hiperparámetros para modelos basados en Seq2Seq

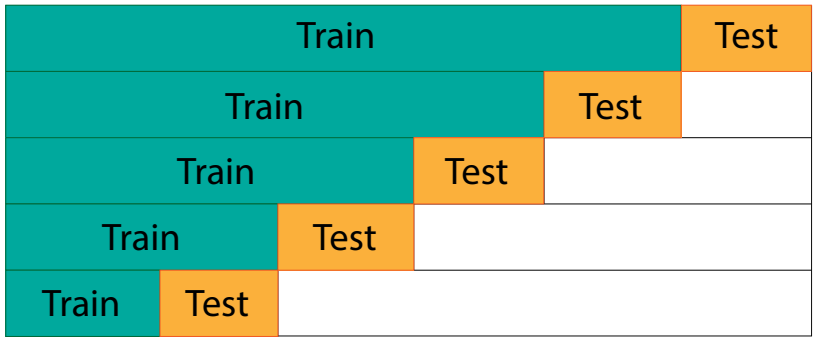
Hiperparámetro	Espacio de búsqueda	Óptimo anual	Óptimo trimestral
Número de pasos pasados	[4, 6, 8, 10, 12, 14, 16, 18, 20]	16	4
Número de células por capa (codificador/decodificador)	[32, 64, 96, 128, 160]	64	64
Ratio de aprendizaje	[10–2, 10–3, 10–4]	10–2	10–3

4.4 Entrenamiento y evaluación de modelos

Para el entrenamiento y la evaluación de los modelos, se emplea la técnica de validación cruzada continua, la cual consiste que a partir del conjunto de datos se generan distintos pares de subconjuntos para entrenamiento y prueba. Los subconjuntos de entrenamiento parten desde el inicio y anteceden a los de prueba ya que son series de tiempo y dependen del orden de sus observaciones (Hyndman & Athanasopoulos, 2021). Para la evaluación se generarán 5 pares de subconjuntos como se ejemplifica en la figura 3. Luego se entrenará un modelo por cada conjunto y se evaluarán sus métricas. El promedio de estas métricas nos servirá para comparar entre los modelos propuestos.

Figura 3

Diagrama de subconjuntos de datos de entrenamiento y prueba usando validación cruzada



Las métricas que se emplearán son el error absoluto medio (MAE), la raíz cuadrada de error cuadrático medio (RMSE) y el error absoluto porcentual medio (MAPE), las cuales están descritas en las ecuaciones 9, 10 y 11. Se consideraron estas métricas basándose de los trabajos de Cook & Smalter Hall (2017), Nguyen y Nguyen (2020) y Kelany et al. (2020).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^*| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2} \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i^*}{y_i}\right| \times 100 \tag{11}$$

5. RESULTADOS Y DISCUSIONES

Se presentan los resultados por cada país en cada subsección, la cual contiene tablas con los promedios de las métricas de desempeño de los modelos en las evaluaciones y se resaltan los valores más bajos. Las gráficas de cajas y bigotes muestran la distribución de las métricas en las evaluaciones individuales. También se grafica la predicción a lo largo de las series, usando los modelos con mejores resultados.

5.1 Argentina

Según los resultados de la tabla 8, en promedio CNN tiene mejor desempeño en datos anuales y MLP para datos trimestrales. En las figuras 4, 5 y 6 observamos que las

métricas individuales no están muy dispersas para los modelos de mejor desempeño a comparación de los modelos restantes. En la figura 7 se pronostica el PBI a lo largo de todo el conjunto de datos y se compara con el estado original del PBI.

Tabla 8

Promedio de métricas en la evaluación de modelos para Argentina

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,039549	0,055836	0,076992	0,048956	0,060864
	RMSE	0,049235	0,068924	0,086598	0,057614	0,072513
	MAPE	8,110362	11,101573	14,114396	9,857729	10,306283
Trimestral	MAE	0,045966	0,051465	0,049192	0,038778	0,058814
	RMSE	0,054223	0,062224	0,059240	0,046067	0,066914
	MAPE	6,813480	7,614213	7,445939	5,817453	8,917134

Figura 4

Distribución de MAE en la evaluación de modelos de Argentina

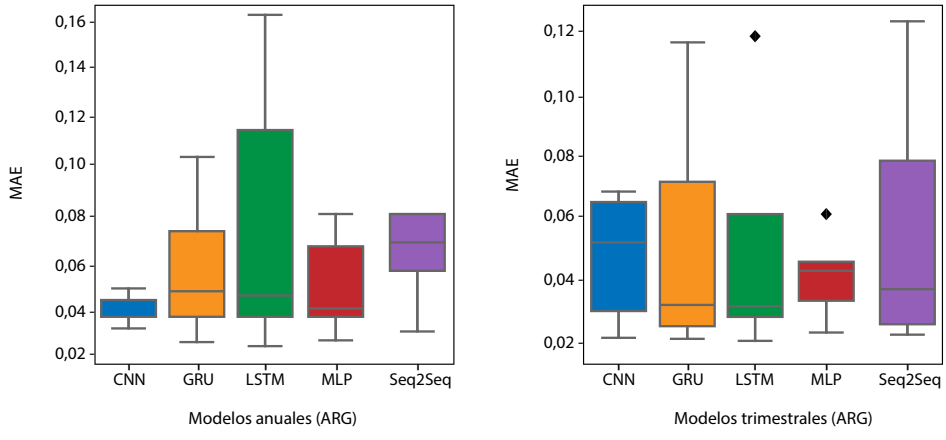


Figura 5

Distribución de RMSE en la evaluación de modelos de Argentina

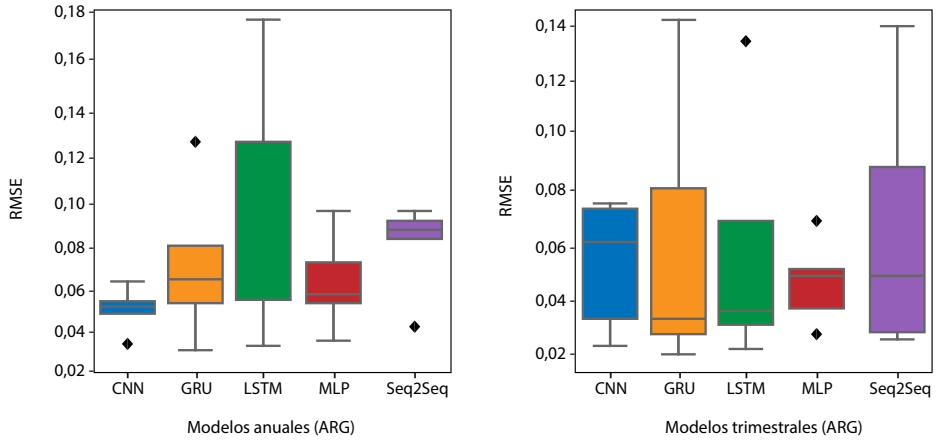


Figura 6

Distribución de MAPE en la evaluación de modelos de Argentina

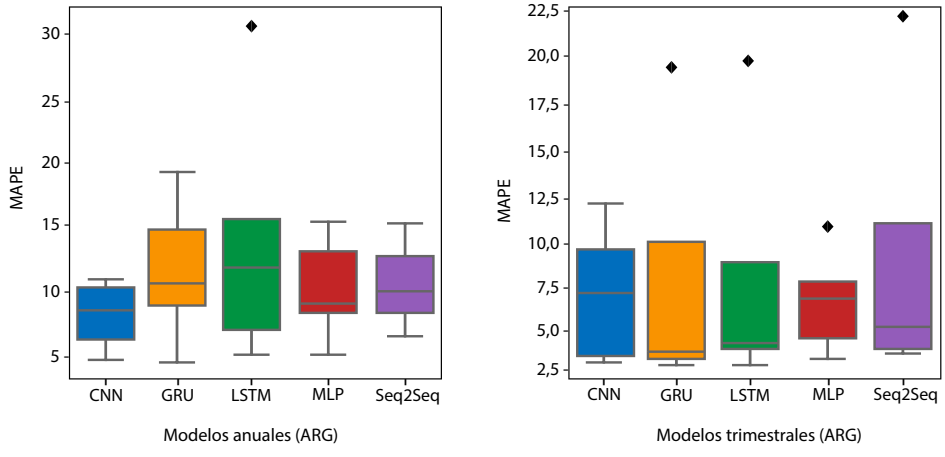
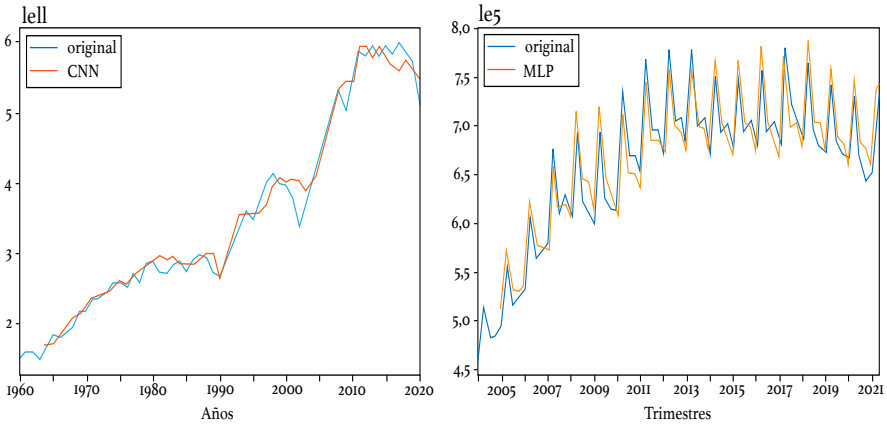


Figura 7

Predicción del PBI usando los modelos con mejores resultados para Argentina



4.2 Brasil

En los resultados de la tabla 9 observamos que, para datos anuales, si bien MLP tiene mejor promedio en MAE, es el modelo Seq2Seq que supera en las otras métricas (RMSE y MAPE). Para datos trimestrales, MLP presenta los mejores promedios. En las figuras 8, 9 y 10 observamos que Seq2Seq tiene menores valores individuales para el caso anual, y lo mismo con MLP en trimestrales. En la figura 11 se pronostica con todo el conjunto de datos y observamos que el modelo Seq2Seq anual predice por debajo de lo esperado en los últimos años.

Tabla 9

Promedio de métricas en la evaluación de modelos para Brasil

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,031627	0,052510	0,037535	0,026192	0,026595
	RMSE	0,035953	0,059284	0,043709	0,032024	0,029517
	MAPE	6,310547	9,748143	7,639060	4,934667	3,642307
Trimestral	MAE	0,030357	0,029789	0,024329	0,023548	0,025053
	RMSE	0,035878	0,036870	0,029941	0,028094	0,030952
	MAPE	4,921294	5,376683	4,354568	4,155495	4,331294

Figura 8

Distribución de MAE en la evaluación de modelos de Brasil

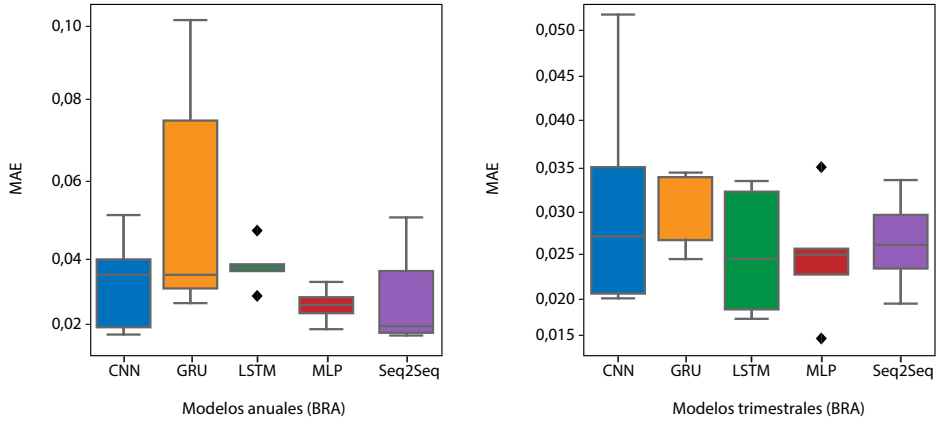


Figura 9

Distribución de RMSE en la evaluación de modelos de Brasil

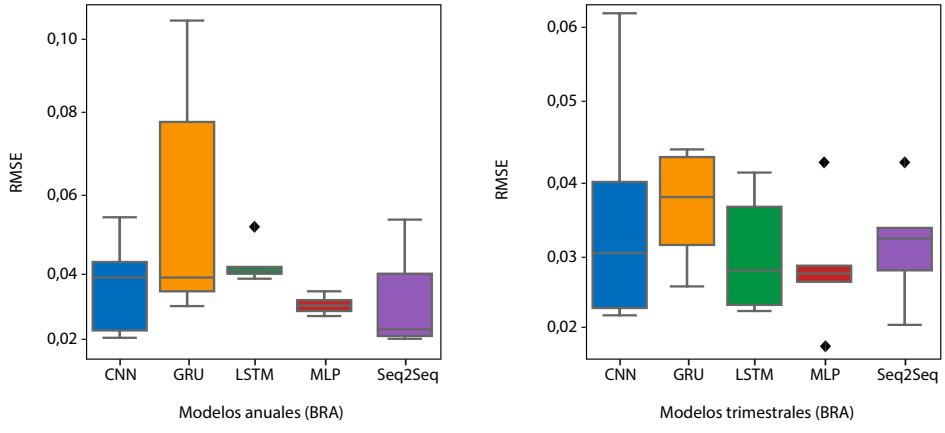


Figura 10

Distribución de MAPE en la evaluación de modelos de Brasil

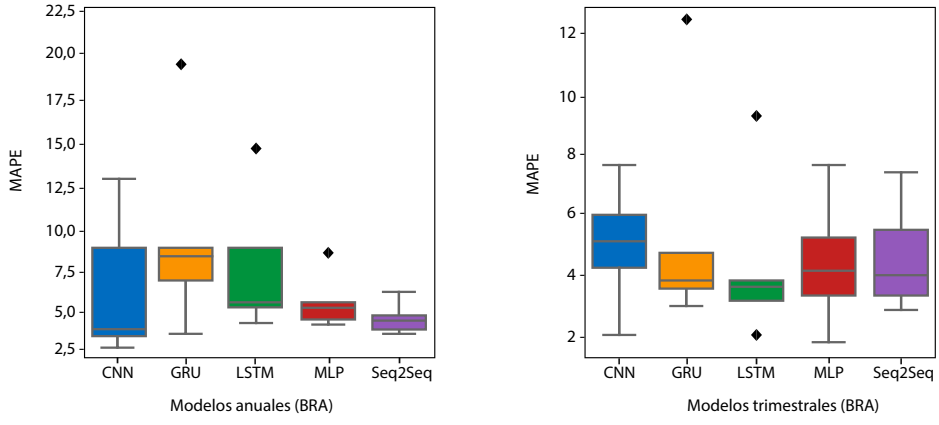
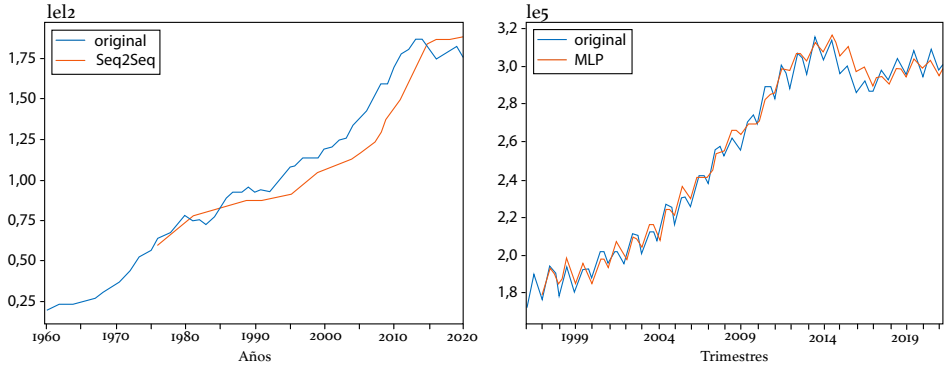


Figura 11

Predicción del PBI usando los modelos con mejores resultados para Brasil



4. Chile

Según los resultados de la tabla 10, en datos anuales, en promedio, MLP tiene mejor desempeño en MAE y RMSE, y Seq2Seq en MAPE. Para datos trimestrales, MLP supera a los restantes. En las figuras 12, 13 y 14 observamos que MLP tiene mejores métricas individuales en los dos casos. La figura 15 pronostica el PBI con los modelos MLP respectivamente.

Tabla 10

Promedio de métricas en la evaluación de modelos para Chile

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,034325	0,033571	0,033604	0,021722	0,024758
	RMSE	0,041179	0,040875	0,039730	0,027051	0,031637
	MAPE	14,892479	12,863911	14,133017	9,772485	7,144571
Trimestral	MAE	0,024589	0,027807	0,031122	0,021260	0,032499
	RMSE	0,031262	0,034194	0,038984	0,026748	0,041087
	MAPE	5,865346	6,656868	7,215769	5,031725	7,081380

Figura 12

Distribución de MAE en la evaluación de modelos de Chile

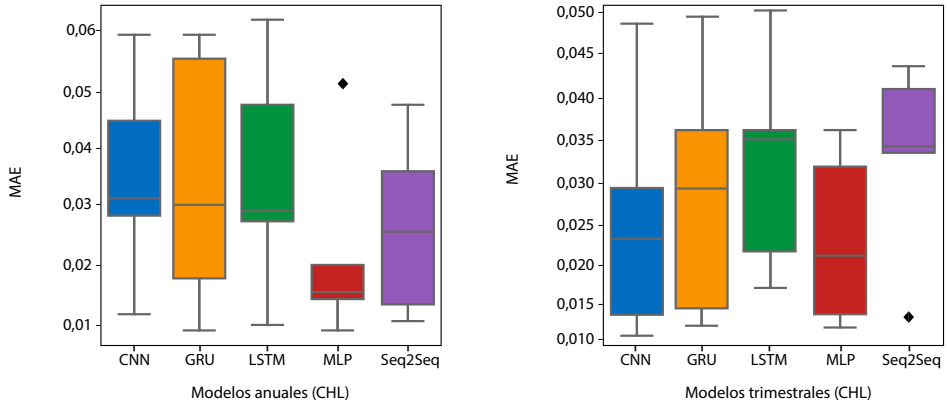


Figura 13

Distribución de RMSE en la evaluación de modelos de Chile

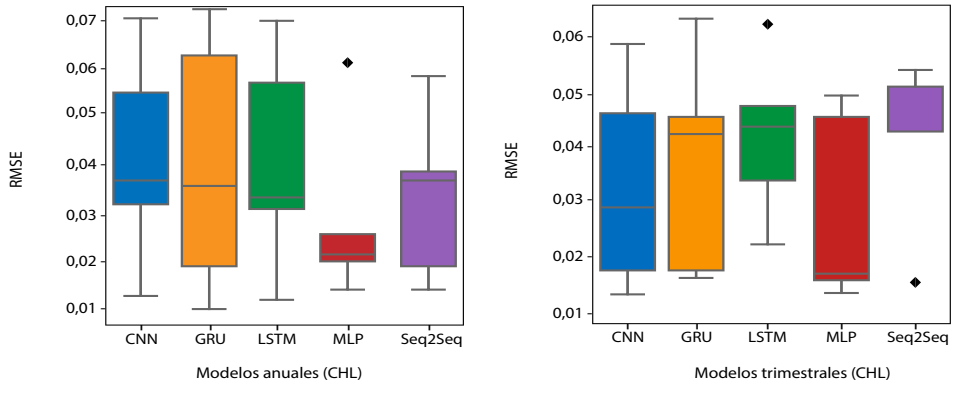


Figura 14

Distribución de MAPE en la evaluación de modelos de Chile

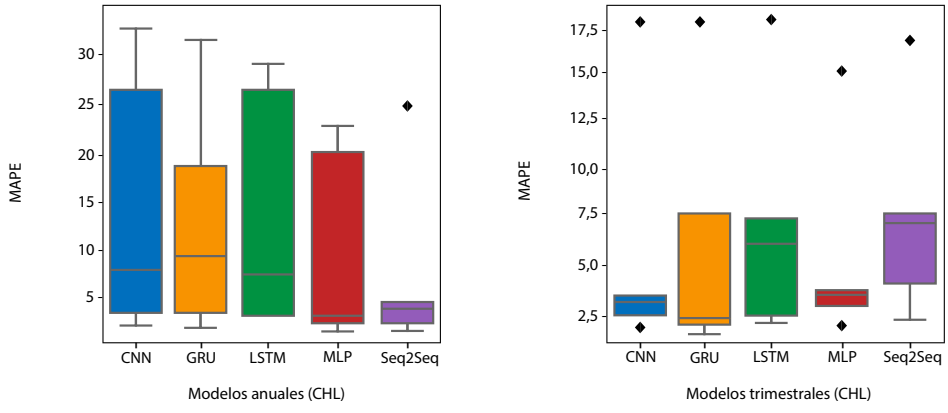
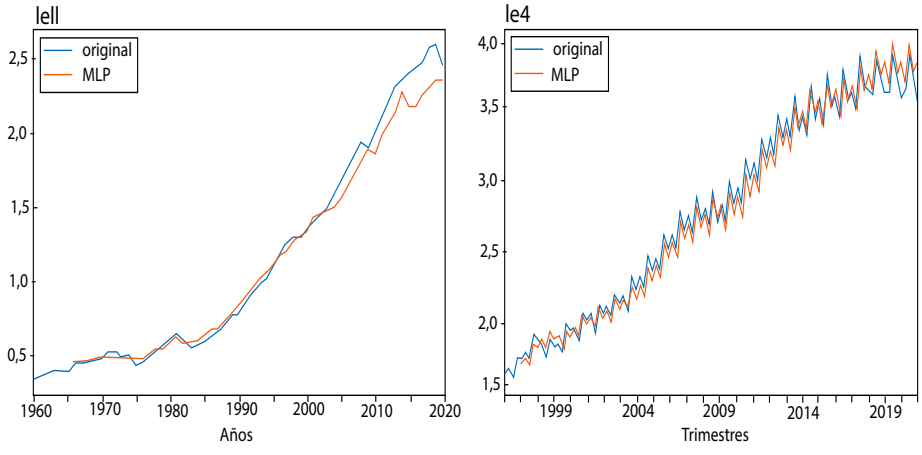


Figura 15

Predicción del PBI usando los modelos con mejores resultados para Chile



4.4 Colombia

En la tabla 11, observamos que, en promedio, para datos anuales, CNN tiene mejor desempeño en MAE y RMSE, y Seq2Seq en MAPE; y para trimestrales, MLP supera a los demás modelos. En las figuras 16, 17 y 18, CNN tiene mejores valores individuales para datos anuales y MLP en trimestrales. En la figura 19 se pronostica el PBI y observamos que CNN se aleja del valor original en los últimos años.

Tabla 11

Promedio de métricas en la evaluación de modelos para Colombia

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,023863	0,036272	0,036205	0,038001	0,026422
	RMSE	0,026892	0,046048	0,041337	0,044638	0,029779
	MAPE	6,390575	9,403895	9,142741	9,585601	5,378279
Trimestral	MAE	0,026070	0,031893	0,040862	0,020612	0,039732
	RMSE	0,031688	0,038621	0,048573	0,024807	0,049222
	MAPE	4,851385	5,814529	7,737369	3,890099	6,512726

Figura 16

Distribución de MAE en la evaluación de modelos de Colombia

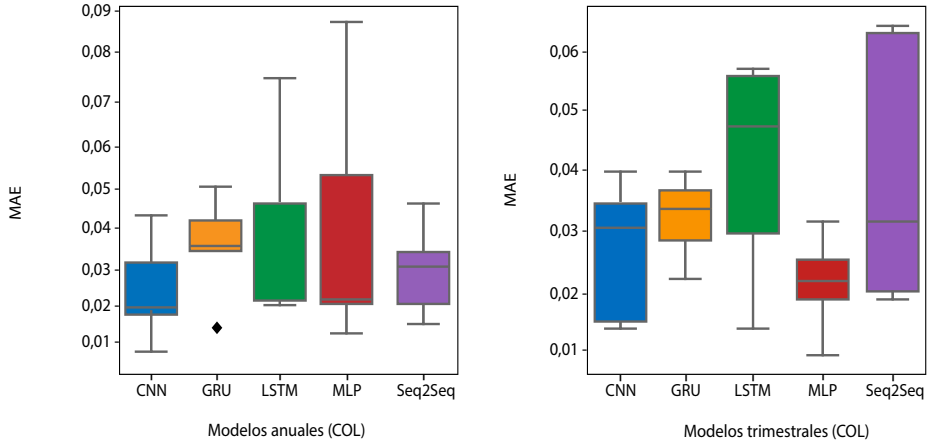


Figura 17

Distribución de RMSE en la evaluación de modelos de Colombia

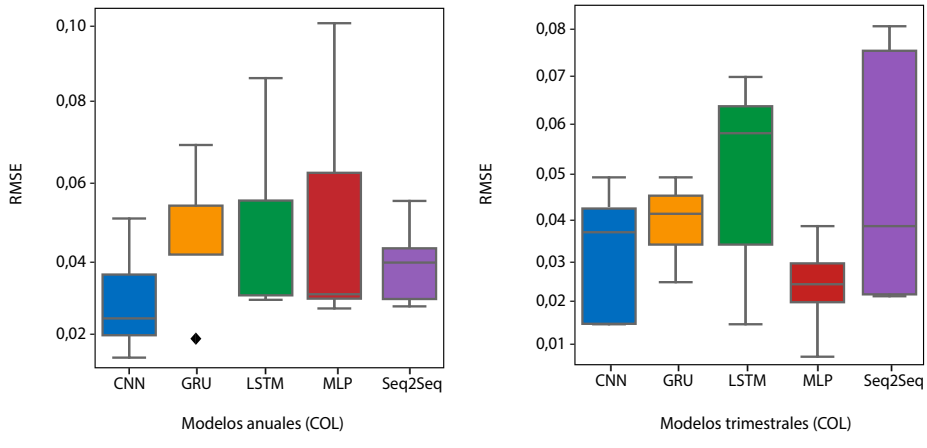


Figura 18

Distribución de MAPE en la evaluación de modelos de Colombia

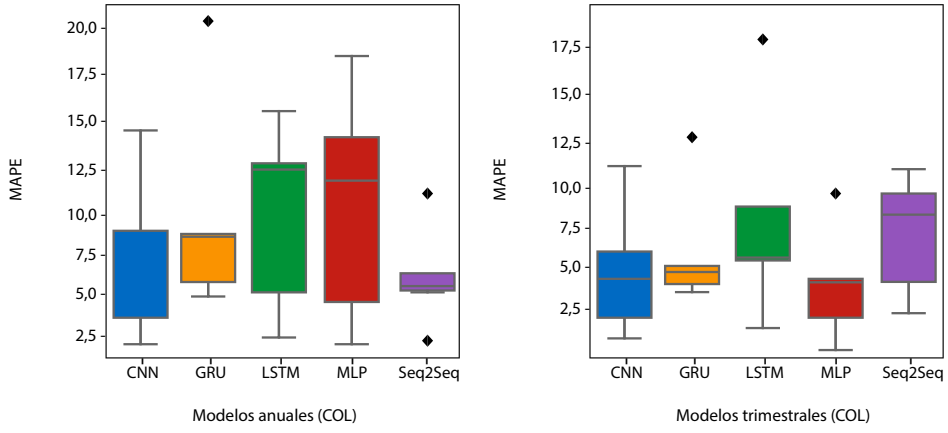
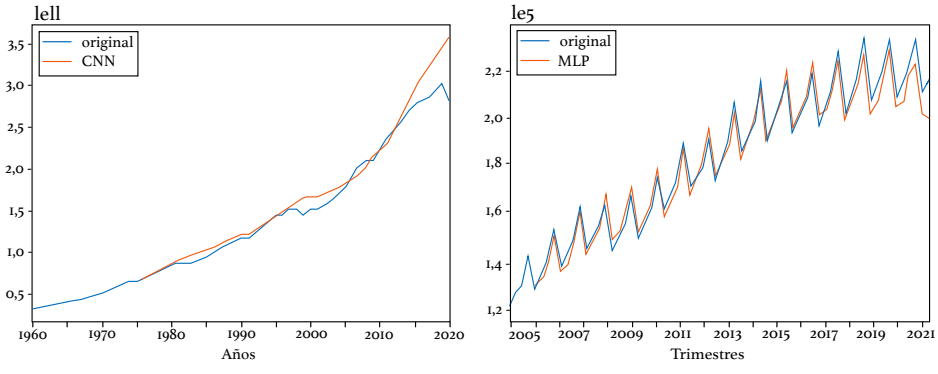


Figura 19

Predicción del PBI usando los modelos con mejores resultados para Colombia



4.5. México

En los resultados de la tabla 12, observamos que el modelo Seq2Seq tiene mejor desempeño en promedio para datos anuales; para trimestrales, el mejor es MLP. En las figuras 20, 21 y 22 se observa que Seq2Seq y MLP tienen los mejores valores individuales respectivamente. En la figura 23 se pronostica el PBI y se realiza la comparación de los resultados a lo largo de todo el conjunto de datosn anuales y trimestrales.

Tabla 12

Promedio de métricas en la evaluación de modelos para México

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,031604	0,044983	0,038439	0,037753	0,029141
	RMSE	0,039263	0,054898	0,045722	0,043099	0,034944
	MAPE	6,539337	8,356517	7,242875	8,492722	4,779532
Trimestral	MAE	0,023512	0,023032	0,024003	0,014631	0,024424
	RMSE	0,028603	0,028353	0,028500	0,017840	0,029421
	MAPE	4,350519	4,232047	4,283476	2,713248	4,448161

Figura 20

Distribución de MAE en la evaluación de modelos de México

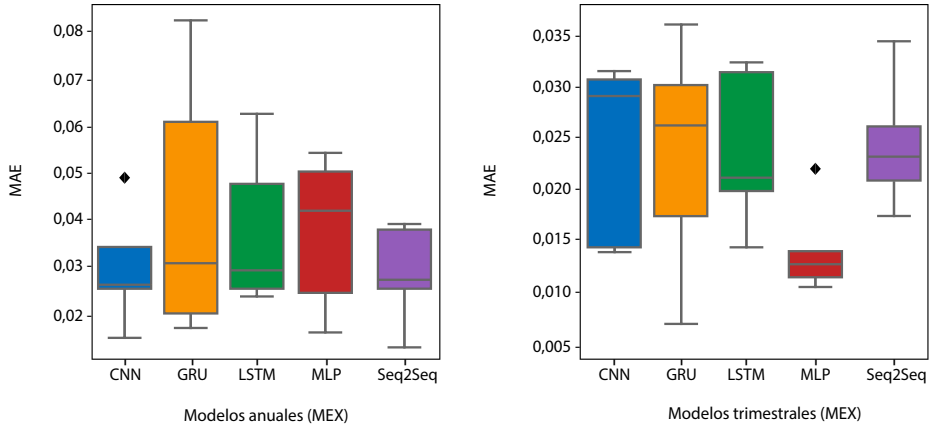


Figura 21

Distribución de RMSE en la evaluación de modelos de México

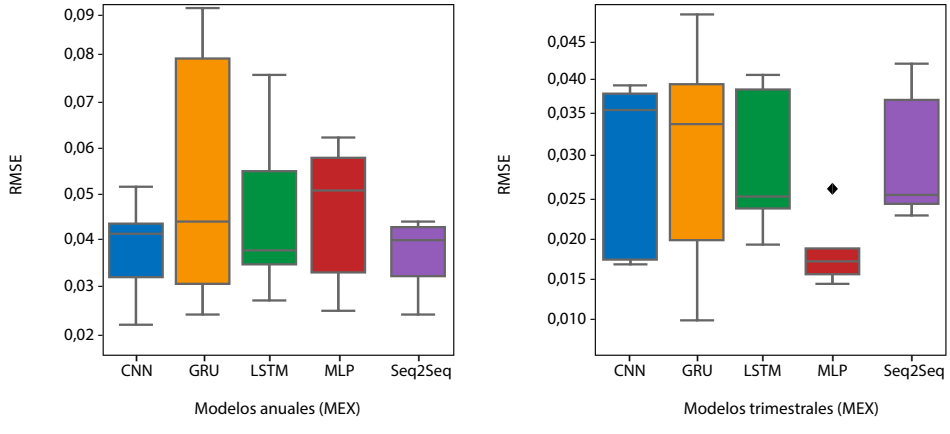


Figura 22

Distribución de MAPE en la evaluación de modelos de México

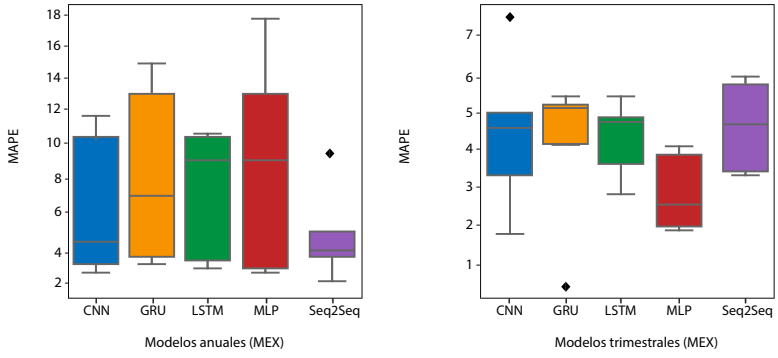
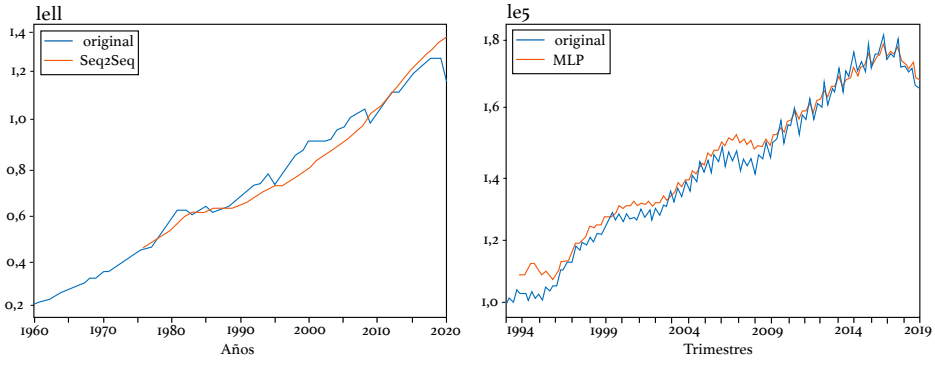


Figura 23

Predicción del PBI usando los modelos con mejores resultados para México



4.6 Perú

Según la tabla 13, en promedio, Seq2Seq tiene mejor desempeño en MAE y MAPE, y MLP en RMSE para datos anuales; mientras que, en trimestrales, MLP supera al resto. En las figuras 24, 25 y 26 se observan que los valores individuales de Seq2Seq no tienen mucha dispersión que los demás en datos anuales, y MLP tiene los valores más bajos en los datos trimestrales del PBI. La figura 27 pronostica el PBI, y en el modelo trimestral se sobreestima el valor predicho en los últimos años.

Tabla 13

Promedio de métricas en la evaluación de modelos para Perú

		CNN	GRU	LSTM	MLP	Seq2Seq
Anual	MAE	0,031455	0,043288	0,053703	0,032230	0,030423
	RMSE	0,039819	0,049165	0,061968	0,038643	0,039474
	MAPE	8,236680	11,280637	13,827532	9,438972	7,486841
Trimestral	MAE	0,025705	0,030373	0,032632	0,021909	0,030306
	RMSE	0,031833	0,036847	0,037814	0,026027	0,035709
	MAPE	4,137657	4,513389	4,620843	3,199139	4,560172

Figura 24

Distribución de MAE en la evaluación de modelos de Perú

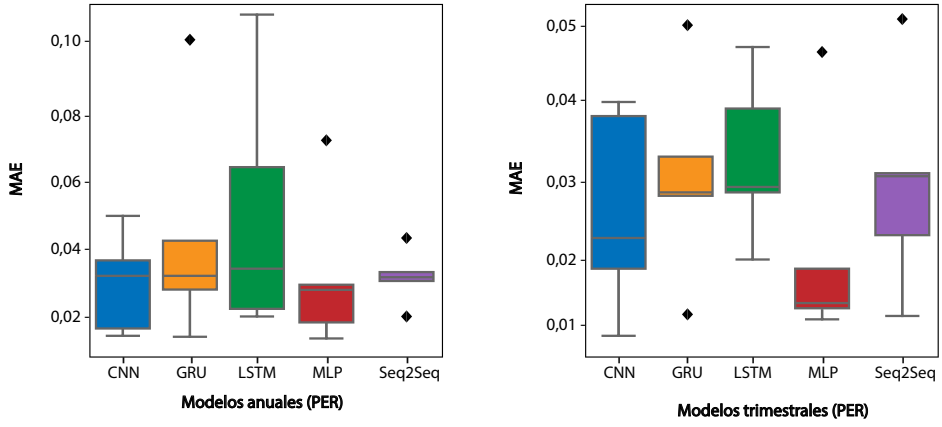


Figura 25

Distribución de RMSE en la evaluación de modelos de Perú

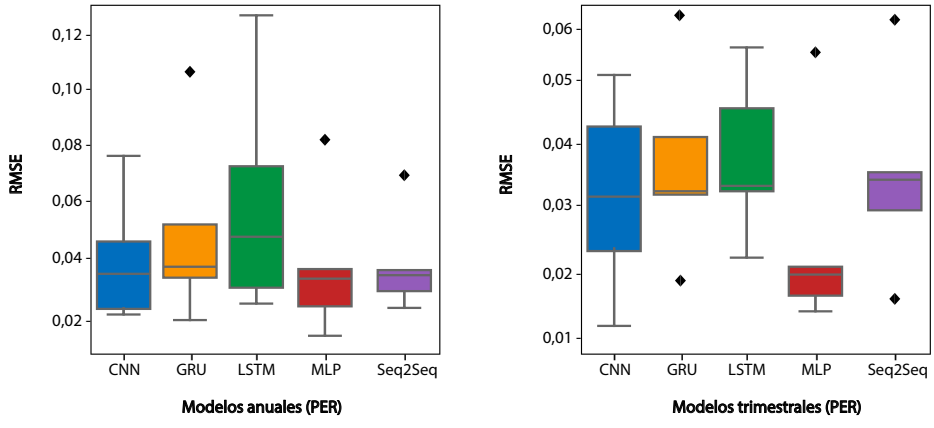


Figura 26

Distribución de MAPE en la evaluación de modelos de Perú

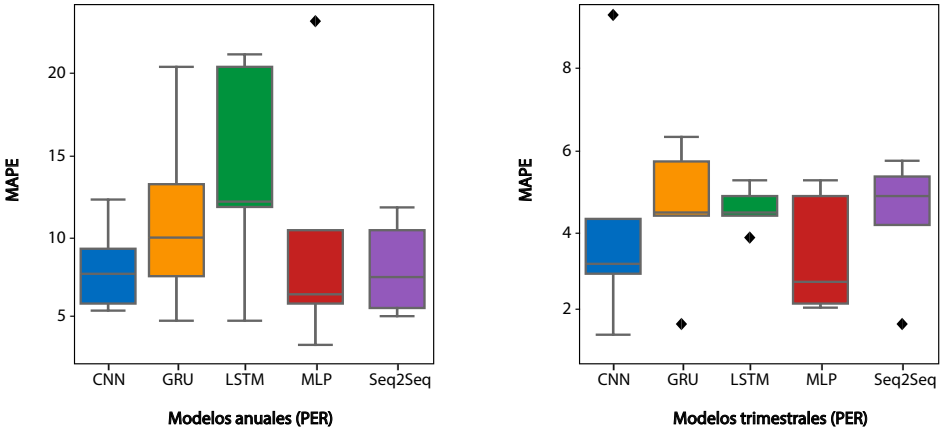
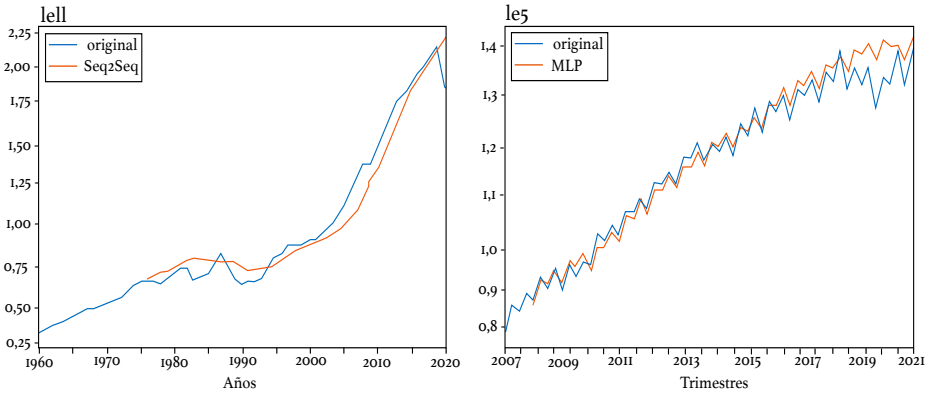


Figura 27

Predicción del PBI usando los modelos con mejores resultados para Perú



4.7 Discusión final

Se halló que los modelos CNN, MLP y Seq2Seq se desempeñan mejor en datos anuales, y que MLP supera a los demás en datos trimestrales. Mientras que, en la predicción del PBI, los modelos trimestrales se aproximan más al original, a excepción del caso de Perú (figura 27) que en el periodo de pandemia de COVID-19 no predice la crisis. Lo mismo sucede con los modelos anuales en las figuras 19, 23 y 27, al sobreestimar el último año que corresponde a la crisis.

5. CONCLUSIONES

Los indicadores macroeconómicos de América Latina tomados de las series anuales del Banco Mundial y trimestrales de la Comisión Económica para América Latina y el Caribe (CEPAL) fueron estudiados en el presente trabajo. El preprocesamiento consiste en la descomposición de las series trimestrales y detección de datos atípicos. Los distintos modelos de *deep learning* usados para pronóstico de las series temporales de los indicadores macroeconómicos fueron el perceptrón multicapa (MLP), red convolucional (CNN), LSTM, GRU y Seq2Seq. Estos modelos se aplicaron a cada país y tipo de conjunto de datos (anual o trimestral). Al evaluar su desempeño con técnica de validación cruzada continua y las métricas MAE, RMSE y MAPE, los modelos tienen diferente desempeño en distintos conjuntos de datos. Para los datos anuales y por país los modelos de mejor desempeño fueron: CNN (Argentina y Colombia), MLP (Chile) y Seq2Seq (Brasil, México y Perú). Y en datos trimestrales fue MLP (los 6 países). Se pronosticó con mayor cercanía el siguiente trimestre en los datos trimestrales. Sin embargo, el pronóstico del siguiente año para los datos anuales se alejó porque el último año observado (2020) corresponde a la pandemia.

Como trabajo a futuro se debería pronosticar también a largo plazo empleando ya no un paso a futuro sino más pasos como salida del modelo. También se debería construir un modelo general el cual se entrene con datos de un grupo de países, y probar con otros países no incluidos en el entrenamiento. Respecto a los datos de entrada, se puede experimentar con series de mayor frecuencia, por ejemplo, mensuales. Además, se pueden incluir más indicadores macroeconómicos no solamente relacionados directamente con el PBI. Esto permitirá experimentar si con más variables mejoraría el pronóstico.

Para el modelamiento, se puede mejorar explorando un mayor espacio de búsqueda de hiperparámetros, incluyendo aquellos que se habían fijado para este trabajo, como, por ejemplo, variar la cantidad de capas de convolución para CNN, la cantidad de capas ocultas en LSTM y GRU; probar otras funciones de activación, de pérdida y optimizadores; buscar el número de épocas óptimas para evitar el sobreajuste.

REFERENCIAS

Banco Mundial. (2021). Banco de datos. Recuperado el 24 de octubre de 2021, de https://databank.bancomundial.org/

Brownlee, J. (2017). *Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

Chollet, F. (2018). *Deep learning with Python*. Manning Publications Co.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3-33.

Comisión Económica para América Latina y el Caribe. (2021). CEPALSTAT: Bases de Datos y Publicaciones Estadísticas. Recuperado el 14 de diciembre de 2021 de https://statistics.cepal.org/portal/cepalstat/dashboard.html

Cook, T., & Smalter Hall, A. (2017). Macroeconomic indicator forecasting with deep neural networks. *The Federal Reserve Bank of Kansas City Research Working Papers*. https://doi.org/10.18651/rwp2017-11

Guimarães, R. R. S. (2022). *Deep learning macroeconomics* [Tesis de maestría, Universidade Federal do Rio Grande do Sul]. http://hdl.handle.net/10183/239533

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3.ª ed.). Elsevier/Morgan Kaufmann.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3.ª ed.), OTexts. Recuperado el 10 de mayo de 2022 de https://OTexts.com/fpp3

Jamieson, K., & Talwalkar, A. (2016). Non-stochastic best arm identification and hyperparameter optimization. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. *Proceedings of Machine Learning Research 51*, 240-248. https://proceedings.mlr.press/v51/jamieson16.html

Jung, J.-K., Patnam, M., & Ter-Martirosyan, A. (2018). An algorithmic crystal ball: Forecasts-based on machine learning. *IMF Working Papers*, *18*(230), 1-33. https://doi.org/10.5089/9781484380635.001

Kelany, O., Aly, S., & Ismail, M. A. (2020, November). Deep learning model for financial time series prediction. *2020 14th International Conference on Innovations in Information Technology (IIT)*, 120-125.

Lazzeri, F. (2020). *Machine learning for time series forecasting with Python*. Wiley.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, *18*(185), 1-52. http://jmlr.org/papers/v18/16-558.html

Mankiw, N. G. (2014). *Macroeconomía*, (8.ª ed.). Antoni Bosch.

Nguyen, H. T., & Nguyen, D. T. (2020). Transfer learning for macroeconomic forecasting. *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, 332-337. https://doi.org/10.1109/NICS51282.2020.9335848

Skansi, S. (2018). *Introduction to deep learning: From logical calculus to artificial intelligence*. Springer.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14)*, *2*, 3104-3112. https://dl.acm.org/doi/10.5555/2969033.2969173

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Education.

Viswanath, S., Saha, M., Mitra, P., & Najundiah R. S. (2019), Deep learning based LSTM and SeqToSeq Models to detect monsoon spells of India. En João M. F. Rodrigues, Pedro J. S. Cardoso, Jânio Monteiro, Roberto Lam, Valeria V. Krzhizhanovskaya, Michael H. Lees, Jack J. Dongarra, Peter M. A. Sloot (Eds.), *Computational Science – ICCS 2019* (Part II, vol. 11537, pp. 204-218). https://doi.org/10.1007/978-3-030-22741-8_15

Zyatkov, N., & Krivorotko, O. (2021). Forecasting recessions in the US Economy using machine learning methods. *2021 17th International Asian School-Seminar "Optimization Problems of Complex Systems (OPCS)"*, 139-146, https://doi.org/10.1109/OPCS53376.2021.9588678