



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA  
INGENIERÍA DE SISTEMAS – OPTIMIZACIÓN FINANCIERA

“PRONÓSTICO DEL RENDIMIENTO DEL ÍNDICE DE PRECIOS Y  
COTIZACIONES DE LA BOLSA MEXICANA DE VALORES MEDIANTE REDES  
NEURONALES PROFUNDAS”

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN INGENIERÍA

PRESENTA:  
JORGE ALBERTO CALVILLO RODRÍGUEZ

TUTOR PRINCIPAL  
DR. FEDERICO HERNÁNDEZ ÁLVAREZ  
FACULTAD DE INGENIERÍA

Ciudad Universitaria, CD. MX. Septiembre 2021



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**JURADO ASIGNADO:**

Presidente: Dra. Rodríguez Vázquez Katya  
Secretario: M.I. Rodríguez Rubio Jorge  
1er. Vocal: Dr. Hernández Álvarez Federico  
2do. Vocal: Dr. Camacho Escoto José Jaime  
3er. Vocal: M.I. García Munguía Fernando Israel

Ciudad Universitaria, CD. MX. Septiembre de 2021

**TUTOR DE TESIS:**

Dr. Hernández Álvarez Federico

---

**FIRMA**



## Abstract

El presente trabajo propone una arquitectura de series de tiempo multivariadas y redes neuronales de aprendizaje profundo para pronosticar el rendimiento del Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores (BMV).

El modelo utiliza como entrada las 20 series accionarias más representativas de la BMV, descompuestas mediante el filtro de Hodrick-Prescott en tendencia, ciclo y residuo. Estas matrices son analizadas por una red neuronal de aprendizaje profundo compuesta por una capa convolucional y dos capas recurrentes con unidades LSTM. La capa convolucional permite atenuar ruido de alta frecuencia y determina correlaciones de corto plazo entre las series, mientras que las capas recurrentes modelan dependencias de mediano y largo plazo.

Una vez entrenados los pesos sinápticos de la red, se generan pronósticos del IPC para los precios de cierre de los próximos dos meses de operación y se compara con los resultados obtenidos mediante un modelo univariado autorregresivo (ARIMA) y un modelo multivariado de Vectores Autorregresivos (VAR). Las métricas de evaluación son el error porcentual absoluto medio (MAPE), el error cuadrático medio (MSE) y el coeficiente  $U$  de Thiel.

Se concluye que la arquitectura de aprendizaje profundo obtiene mejores resultados, tanto técnicos como financieros, al ser capaz de extraer dependencias no lineales de corto y largo plazo en series de tiempo multivariadas, facilitando el procesamiento y análisis de grandes volúmenes de información.

**Palabras Clave:** series de tiempo multivariadas, aprendizaje profundo, redes neuronales convolucionales, redes neuronales recurrentes.



## Resumen

El pronóstico de índices financieros es un problema complejo debido a que el mercado bursátil está influenciado por diversos factores subyacentes e interacciones que están en constante cambio y que, en muchos casos, no son observables.

La dinámica del mercado de valores se asocia con cambios en la macroeconomía y en decisiones individuales de inversión. Por lo tanto, sus índices no solo reflejan el movimiento de la economía financiera, sino que también sugieren modificaciones en industrias sustanciales, lo que influye en gran medida en la trayectoria de la economía nacional.

Por otra parte, existe una creciente disponibilidad de datos financieros, pero su análisis requiere de una ardua labor y sus valores son difíciles de predecir con precisión a causa de las tendencias no lineales y al ruido presente en la series.

Esto crea la necesidad de definir técnicas robustas y eficientes, capaces de extraer información en grandes volúmenes de datos, inferir patrones subyacentes en los movimientos de las diferentes series que componen los índices y modelar las interdependencias de mediano y largo plazo.

En años recientes, los modelos de aprendizaje profundo han revolucionado diferentes ámbitos de la ciencia y la industria, convirtiéndose en un área de investigación extremadamente activa que prepara el camino para la inteligencia artificial moderna. Cada vez que se utilizan aplicaciones como traductores en tiempo real, de reconocimiento facial o vehículos autónomos, se está ante algoritmos de aprendizaje profundo.

El ámbito financiero no es la excepción. Cada vez existen más aplicaciones de las técnicas de aprendizaje profundo en áreas como administración de portafolios (Hu y Lin, 2019), otorgamiento de préstamos y seguros (Ablad, Frikh, y Ouhbi, 2020), detección y prevención de fraudes (Babu y Pratap, 2020), entre otros.

El presente trabajo propone una arquitectura de redes neuronales de aprendizaje profundo para pronosticar el rendimiento del Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores (BMV), el índice bursátil más importante del mercado financiero en México.

Para realizar los pronósticos se aplica un enfoque de series de tiempo multivariadas que toma como datos de entrada las 20 series accionarias más representativas del mercado mexicano y que presentan una mayor correlación con el IPC. Se utilizan los precios de cierre de 753 días de operación (3 años) y se realizan 6 pronósticos utilizando 452 días

(85 %) para entrenamiento de la red, 42 días (7.5 %) para optimizar los hiperparámetros del modelo y 42 días (7.5 %) para generar pronósticos de prueba.

En la etapa de preprocesamiento, los datos son suavizados utilizando el filtro de Hodrick-Prescott para extraer la tendencia, el ciclo y el residuo para cada serie, generando tres matrices de datos que servirán como entrada para el modelo.

La red neuronal profunda cuenta con una capa convolucional, que extrae interdependencias de corto plazo entre las series, y con dos capas recurrentes con unidades LSTM apiladas que modelan las interdependencias de mediano y largo plazo.

El modelo planteado se prueba empíricamente en el conjunto de datos antes mencionado y se compara con los pronósticos generados mediante un modelo univariado autorregresivo (ARIMA) y un modelo multivariado de Vectores Autorregresivos (VAR), utilizando como métricas el error porcentual absoluto medio (MAPE), el error cuadrático medio (MSE) y el coeficiente  $U$  de Thiel. Los resultados obtenidos muestran que el modelo propuesto logra un mejor rendimiento de pronóstico y facilita el manejo de grandes volúmenes de información financiera.

La tesis está organizada como sigue: el capítulo 1 presenta el problema de generar pronósticos, particularmente en el ámbito económico y financiero, justificando el desarrollo de modelos matemáticos para este propósito. En el capítulo 2 se describen los antecedentes sobre la generación de pronósticos en mercados financieros con aprendizaje automático, las técnicas usadas y sus principales resultados.

En el capítulo 3 se exponen los conceptos teóricos en los que se sustenta el modelo de aprendizaje profundo propuesto. En el capítulo 4 se presentan los detalles de la arquitectura empleada, así como el análisis exploratorio de los datos utilizados.

En el capítulo 5 se describen los experimentos realizados, los resultados obtenidos y la evaluación de los mismos. Finalmente, en el capítulo 6 se plantean las conclusiones y el trabajo a futuro.



# Índice general

<b>Abstract</b>	<b>I</b>
<b>Resumen</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Justificación . . . . .	1
1.1.1. Pronósticos . . . . .	1
1.1.2. Mercado de Capitales . . . . .	3
1.2. Planteamiento del Problema . . . . .	5
1.3. Hipótesis . . . . .	5
1.4. Objetivos . . . . .	5
1.4.1. Objetivo General . . . . .	5
1.4.2. Objetivos Particulares . . . . .	6
<b>2. Estado del Arte</b>	<b>7</b>
2.1. Pronósticos en Mercados Financieros . . . . .	7
2.2. Técnicas Tradicionales de Pronósticos . . . . .	8
2.3. Pronósticos en Mercados Financieros con Redes Neuronales Artificiales . . . . .	9
2.4. Pronósticos en Mercados Financieros con Aprendizaje Profundo . . . . .	11
2.5. Pronósticos de Mercados Financieros en México con Aprendizaje Automático . . . . .	13
<b>3. Marco Teórico</b>	<b>15</b>
3.1. Inteligencia Artificial . . . . .	15
3.2. Redes Neuronales Artificiales . . . . .	16

3.2.1.	El Perceptrón . . . . .	16
3.2.2.	Funciones de Activación . . . . .	18
3.2.3.	Perceptrón Multicapa . . . . .	19
3.2.4.	Algoritmo de Aprendizaje . . . . .	21
3.3.	Redes Neuronales Profundas . . . . .	22
3.3.1.	Redes Neuronales Recurrentes . . . . .	24
3.3.2.	Redes de Memoria a Corto y Largo Plazo (LSTM) . . . . .	26
3.3.3.	Redes Neuronales Convolucionales . . . . .	30
<b>4.</b>	<b>Metodología</b>	<b>33</b>
4.1.	Datos . . . . .	33
4.1.1.	Selección de Series Características . . . . .	34
4.2.	Arquitectura del Modelo . . . . .	36
4.2.1.	Descomposición en tendencia, ciclo y residuo . . . . .	36
4.2.2.	Capa Convolutiva . . . . .	39
4.2.3.	Capas Recurrentes LSTM . . . . .	41
<b>5.</b>	<b>Resultados</b>	<b>44</b>
5.1.	Experimentos . . . . .	44
5.2.	Evaluación de la precisión predictiva . . . . .	46
5.3.	Evaluación de los rendimientos. . . . .	49
<b>6.</b>	<b>Conclusiones y Trabajo Futuro</b>	<b>51</b>
	<b>Referencias</b>	<b>53</b>
	<b>Apéndices</b>	<b>59</b>
	<b>A. Muestra del IPC</b>	<b>59</b>
	<b>B. Matriz de Correlación</b>	<b>61</b>

# Índice de cuadros

4.1. Muestra del IPC . . . . .	37
5.1. Periodos . . . . .	45
5.2. Precisión Predictiva: MAPE . . . . .	48
5.3. Precisión Predictiva: MSE . . . . .	48
5.4. Precisión Predictiva: RMSE . . . . .	48
5.5. Precisión Predictiva: Thiel U . . . . .	48
5.6. Rendimientos . . . . .	50
A.1. Muestra del IPC . . . . .	60



# Índice de figuras

3.1. Perceptrón . . . . .	17
3.2. Datos no linealmente separables . . . . .	18
3.3. Clasificación con funciones de activación lineal y no lineal . . . . .	19
3.4. Red neuronal densa . . . . .	20
3.5. Ubicación del aprendizaje profundo . . . . .	23
3.6. Error en clasificación de imágenes . . . . .	24
3.7. Red neuronal recurrente . . . . .	25
3.8. Red neuronal recurrente con función de activación <i>tanh</i> . . . . .	26
3.9. RNN con dependencias de corto plazo . . . . .	26
3.10. RNN con dependencias de largo plazo . . . . .	27
3.11. Secuencia de unidades LSTM . . . . .	27
3.12. Capa de la puerta de olvido . . . . .	28
3.13. Capa de la puerta de entrada . . . . .	29
3.14. Actualización del estado $C_t$ . . . . .	29
3.15. Salida de la unidad LSTM . . . . .	30
3.16. Ejemplo operación convolución en 2D . . . . .	32
4.1. Correlación IPC y series accionarias . . . . .	35
4.2. Series accionarias con mayor información . . . . .	36
4.3. Arquitectura propuesta . . . . .	38
4.4. Descomposición de la serie en tendencia, ciclo y residuo . . . . .	39
4.5. Matriz de tendencias . . . . .	40
4.6. Matriz de ciclos . . . . .	40

4.7. Matriz de residuos . . . . .	40
4.8. Arquitectura de aprendizaje profundo . . . . .	43
5.1. Segmentación de datos . . . . .	45
5.2. Precio de cierre ajustado del IPC en 2019 . . . . .	45
5.3. Real y pronóstico CNN-LSTM para los periodos de 2019 . . . . .	46
5.4. Pronósticos para los periodos de 2019 . . . . .	47
6.1. Ciclo real y pronóstico CNN-LSTM para 2019 . . . . .	52
B.1. Matriz de Correlación . . . . .	61

# Capítulo 1

## Introducción

### 1.1. Justificación

#### 1.1.1. Pronósticos

Desde los inicios de la ciencia, la humanidad ha intentado realizar predicciones y desarrollar tecnología que le permita vaticinar eventos futuros, tales como determinar el movimiento y posición de los astros en el cielo o calcular la temporada de cosecha.

Todos los días cada decisión que tomamos, por insignificante que parezca, se fundamenta en la predicción de los resultados esperados y en ciertas condiciones que, suponemos, estarán presentes para su realización. Las personas toman decisiones basadas en la creencia, explícita o implícita, de que una acción conducirá a recompensas mayores que las que acarrearía una acción alternativa.

Sin embargo, lo que permite distinguir a los tiempos modernos del pasado distante es la noción de que el futuro no es un simple capricho de los dioses y que la humanidad no es pasiva ante un supuesto destino (Bernstein, 1996), sino más bien activa, lo que posibilita tanto incidir en los hechos como predecirlos.

Podemos definir el concepto de pronóstico como el cálculo de eventos futuros dados diferentes escenarios, cimentado en datos y supuesto sobre factores que se consideran influyentes y que ayudan a comprender lo que se espera que ocurra. Los pronósticos pueden ser extrapolaciones de tendencias simples, pero, de manera más específica, implican

modelos estadísticos y análisis cuantitativos.

Métodos matemáticos como el análisis de series de tiempo o las redes neuronales artificiales ofrecen un medio racional para realizar pronósticos. Sin embargo, una condición previa para llevarlos a cabo es la disponibilidad de una secuencia de valores del pasado observado cuya clasificación temporal permite la deducción de los atributos necesarios para fines de pronóstico (Moller y Reute, 2007). En otras palabras, un pronóstico es posible debido a que determinadas dependencias temporales pueden deducirse de una secuencia significativa de los valores observados en el pasado.

Aunque los pronósticos se utilizan ampliamente en diferentes disciplinas, en economía y finanzas son herramientas fundamentales para la toma de decisiones y la planificación al más alto nivel ejecutivo (Kharin, 2013). Algunos ejemplos de dichas aplicaciones son:

- Planificación y regulación estatal de las economías nacionales.
- Pronóstico del PIB, la inflación y otros indicadores macroeconómicos.
- Pronóstico de las tasas de interés futuras.
- Calificación crediticia en la banca.
- Pronóstico de las tasas del mercado de valores y las tasas de cambio en divisas.
- Previsión de futuras reclamaciones sobre pólizas de seguro.
- Pronóstico de ventas y proyección de gastos.
- Pronóstico de la demanda total de productos.

Los pronósticos son cruciales para toda actividad económica; ya sean grandes corporaciones, pequeñas empresas o tomadores de decisiones en el sector público, los pronósticos son una herramienta fundamental que permite a los agentes económicos evaluar las condiciones actuales y estimar las condiciones futuras. Estos pronósticos son una guía para la toma de decisiones informada y, por lo tanto, un componente esencial de una buena gestión. Adicionalmente, los pronósticos generan confianza y seguridad en una economía, haciendo la toma de decisiones más eficiente.



Cuando el *Homo Economicus* analiza las previsiones económicas y financieras o los proyectos de presupuesto, forma las expectativas que determinan su comportamiento, independientemente de la calidad, precisión y fiabilidad de estas predicciones.

La imprecisión de los pronósticos (o su ausencia) puede conducir a gestiones que no reconocen ni gestionan los riesgos, que intervienen de forma tardía, desaprovechando oportunidades de inversión, perdiendo participación en el mercado, reduciendo las ganancias e incluso llegando a la bancarrota. En el sector público, una mala predicción puede significar que los proyectos tengan costos adicionales, que se completen más tarde o produzcan menos beneficios. Los pronósticos macroeconómicos inexactos, ya sea que subestimen o sobre estimen, conducen a la economía por un mal camino. En su estudio, Tetlock (2015) investiga la precisión de los pronósticos del ámbito político y, después de recopilar y analizar 28,000 predicciones de 284 expertos, concluye que los pronósticos promedio de los expertos eran solo un poco más precisos que predicciones generadas de forma aleatoria.

### 1.1.2. Mercado de Capitales

El mercado de capitales es una parte del mercado financiero que realiza funciones financieras y económicas. En términos de la economía, el mercado de capitales es un lugar de encuentro para asignar fondos de manera eficiente entre los agentes económicos que tienen fondos excedentes (el lado de la oferta) y aquellos que necesitan fondos (el lado de la demanda), asumiendo cierto riesgo y rendimiento, lo que nos conduce a su función financiera.

Las actividades del mercado de capitales tienen como propósito obtener rendimientos para quienes asignan fondos de acuerdo con el tipo de inversiones que han elegido y para los proyectos productivos que se financian. Debido a que el desarrollo financiero promueve a su vez el desarrollo de la economía, los gobiernos de todo el mundo siempre se han comprometido a desarrollar las finanzas nacionales, especialmente los mercados financieros y las instituciones financieras.

Una de las principales características del mercado de capitales es que influye en la percepción pública y da forma al panorama económico. Sus indicadores pueden infundir

confianza en las empresas para expandir sus operaciones y asumir riesgos. En estos casos, las empresas contratan a más trabajadores, mejoran la tasa de empleo, generando a su vez más ingreso disponible. Las caídas en los indicadores del mercado señalan lo contrario: las empresas se preocupan por cómo financiar sus operaciones, aumentan los despidos y los consumidores no gastan la misma proporción de su ingreso disponible.

Un factor que ha incrementado sustancialmente la complejidad del entorno económico es el fenómeno de la globalización. Los mercados financieros internacionales se han expandido y el volumen de transacciones en los mercados de capital y de divisas han aumentado notablemente desde finales del siglo pasado (Okina y Shiratsuka, 1999). El progreso en las tecnologías de la información y en las telecomunicaciones, la liberalización en los mercados de capitales y el desarrollo de nuevos instrumentos financieros han estimulado aún más los flujos internacionales de capital, lo que ha producido también un incremento del efecto de contagio en momentos de crisis financieras, “en esencia, los brotes de crisis financieras son siempre el resultado de la transmisión de riesgos y la acumulación de riesgos en los mercados financieros” (Wu y Duan, 2019).

Por otra parte, puntos de vista más modernos para estudiar los mercados financieros parten de la hipótesis de que son sistemas complejos en evolución. Utilizan el enfoque de la mecánica estadística y modelan a los agentes económicos como si fuesen una partícula de gas (Ferrari, 2005), dando lugar a toda una nueva disciplina: la econofísica.

Más recientemente, Cassese (2020) analiza el trabajo de Arrow (1964) sobre la completitud de los mercados y concluye que la complejidad de los sistemas económicos modernos está creciendo tan rápido que hace que la completitud del mercado sea un proceso continuo. Por lo tanto, en un mundo complejo, los mercados financieros se comportan siempre como incompletos.

Otros estudios (Kuhlmann, 2014) han observado que, sin causas exógenas, los precios del mercado financiero se desvían significativamente del comportamiento gaussiano de las variables aleatorias, indicando fuertes interacciones no lineales de sus subunidades (agentes). Esto provoca que hacer previsiones del rumbo de los mercados financieros sea particularmente difícil: incluso si se sabe lo que va a suceder, es posible que no se pueda

anticipar cómo responderán los agentes y cómo afectarán sus reacciones al comportamiento general del mercado.

## 1.2. Planteamiento del Problema

Hoy más que nunca, los agentes económicos están obligados a operar en ambientes complejos, de alta incertidumbre y cambios vertiginosos. El que las previsiones puedan ser inexactas crea un serio dilema para los responsables de la toma de decisiones y hace indispensable el desarrollo de modelos matemáticos que faciliten la generación de pronósticos que garanticen en cierta medida su confiabilidad.

De este modo, cabe preguntarse: ¿cuáles son esos modelos matemáticos que permiten desarrollar pronósticos confiables en escenarios como los que plantean los mercados financieros actuales?

## 1.3. Hipótesis

La hipótesis de este trabajo consiste en afirmar que las técnicas de aprendizaje profundo aplicadas a la predicción de índices financieros, facilitan la extracción de conocimiento en grandes volúmenes de información, permiten modelar interdependencias no lineales entre las series de tiempo y producen predicciones más precisas que las derivadas de los métodos econométricos tradicionales.

## 1.4. Objetivos

### 1.4.1. Objetivo General

Pronosticar el precio de cierre y el rendimiento del Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores (BMV)

### 1.4.2. Objetivos Particulares

- Construir una arquitectura de aprendizaje profundo compuesto por una red neuronal convolucional que permita modelar las interacciones de corto plazo entre las series de datos y redes neuronales recurrentes con unidades LSTM que modelen las interacciones y dependencias de mediano y largo plazo.
- Evaluar el desempeño de la arquitectura de aprendizaje profundo en comparación con enfoques econométricos tradicionales, como los modelos univariados autorregresivos (ARIMA) y enfoques aplicados a series de tiempo multivariadas como los vectores autorregresivos (VAR).

# Capítulo 2

## Estado del Arte

### 2.1. Pronósticos en Mercados Financieros

El primer estudio sobre el comportamiento de la trayectoria futura del mercado de valores fue realizado por Bachelier (1900), quien concluyó que el movimiento del precio de las acciones se comporta a la manera de una caminata aleatoria. Esta característica no determinista del precio de los activos financieros fue probada empíricamente por Fama (1965) y apoyada por el trabajo de Cootner (1968).

Estas investigaciones se basaron en la Hipótesis del Mercado Eficiente (HME) que supone que el precio del activo refleja toda la información disponible y, por lo tanto, el movimiento futuro del precio de los activos es independiente de la información pasada y presente (Fama, 1970). El nivel de información puede dividirse en tres grupos: la información pasada del precio de las acciones, la información pública y toda la información existente (incluida la información privada). Cada conjunto de información determina un nivel de eficiencia del mercado: la forma de eficiencia débil, la forma semi-eficiente y la forma completamente eficiente.

En particular, la prueba de la eficiencia débil tiene como objetivo investigar si la información contenida en los precios posteriores de los activos refleja completamente los precios actuales. Si se prueba esta hipótesis, significa que los cambios en los precios anteriores no están correlacionados con los precios actuales de los valores, por lo que no se puede usar esta información para predecir los precios y rendimientos de los activos finan-

cieros y, en consecuencia, los instrumentos de análisis diseñados para predecir los precios futuros de las acciones, así como para proporcionar señales para comprar, mantener o vender, no tienen sentido. Por el contrario, si esta hipótesis no se prueba, significa que los inversores pueden hacer predicciones con técnicas que utilizan la información histórica de los precios, tal como el análisis de series de tiempo (Hawaldar, Rohit, y Pinto, 2017).

Los opositores a la HME señalan que existen inversores que constantemente generan retornos de inversión que superan el rendimiento del mercado general. De acuerdo a la HME eso debería ser posible únicamente por el azar, pero eso no explicaría por qué las mismas personas superan los rendimientos del mercado por un amplio margen de manera repetida en un considerable lapso de tiempo. Esta idea ha dado lugar a estudios que intentan refutar experimentalmente la HME (Georgieva, 2016).

La evidencia empírica ha demostrado que los mercados de valores son, hasta cierto punto, predecibles. La realidad económica ha permitido observar que existen retrasos entre la disponibilidad de información y la reacción del mercado. Además, es posible encontrar que el grado de eficiencia entre varios mercados (o entre activos de un mismo mercado) puede diferir. Por ejemplo, en las economías menos desarrolladas la información puede ser más difícil de descubrir o de incorporar a los modelos predictivos, dando un margen a aquellos inversores mejor preparados para obtener rendimientos extraordinarios.

## 2.2. Técnicas Tradicionales de Pronósticos

Entre los modelos tradicionales para hacer pronósticos en mercados financieros, se encuentran modelos estadísticos paramétricos (Box, Jenkins, y Reinsel, 2008) como el promedio móvil autorregresivo (ARMA autoregressive moving average), el promedio móvil integrado autorregresivo (ARIMA autoregressive integrated moving average) y los vectores autorregresivos (VAR vector autoregression), métodos adecuados para sistemas discretos con datos muestreados a intervalos de tiempo igualmente espaciados (Cowpertwait y Metcalfe, 2011). Sin embargo, estos modelos solo son capaces de captar características y dependencias lineales en los datos que, en el caso de las variables económicas, se generan de forma no sincrónica y con diferentes grados de retraso (Modugno, 2011). Esto ha dado

lugar a la necesidad de abordar el análisis de los mercados financieros con modelos no paramétricos y herramientas no lineales.

## 2.3. Pronósticos en Mercados Financieros con Redes Neuronales Artificiales

En años recientes, las técnicas de aprendizaje automatizado y las redes neuronales artificiales han cobrado una gran popularidad en diferentes áreas del conocimiento y el campo de los pronósticos financieros no es la excepción. De acuerdo con López de Prado, “en un futuro cercano, el aprendizaje automático (machine learning) dominará las finanzas, la ciencia reducirá las conjeturas y la inversión ya no significará juegos de azar” (López de Prado, 2018), aunque no omite señalar que para que esto suceda, hace falta mucho trabajo de investigación, ya que vencer la sabiduría de las multitudes en el mercado de valores, es sin duda más difícil que reconocer rostros o clasificar imágenes de animales.

De acuerdo con Vanstone y Finnie (2009), las redes neuronales son una técnica que muestra una alta tolerancia para funcionar de manera adecuada en entornos de datos altamente ruidosos, características que las hacen una herramienta idónea para abordar problemas del ámbito financiero.

El compendio realizado por Atsalakis y Valavanis (2009), analizando más de 100 artículos científicos relacionados con pronósticos del mercado de valores, concluye que las redes neuronales artificiales y los modelos neurodifusos, son apropiados para realizar pronósticos en el mercado de valores. Los resultados generales demuestran que estos modelos superan a los modelos convencionales en la mayoría de los casos, tanto como sistemas de negociación o como mecanismos de pronóstico. Sin embargo, apunta las dificultades que surgen al definir la arquitectura del modelo, como las capas ocultas o el número de neuronas, cuestiones que hasta el momento solo pueden resolverse mediante procedimientos de prueba y error.

Adicionalmente, la predicción de series temporales financieras por redes neuronales requiere una preparación especial de datos. Wang *et al.* (2011) desarrollaron un algorit-

mo para predecir los precios de cierre del Índice Compuesto de Shanghai (SCI Shanghai Composite Index), proponiendo una red neuronal de retro-propagación, pero suavizando el ruido mediante una función de transformada *wavelet*. En este método, primero se descomponen los datos originales en múltiples capas mediante *wavelets*; cada capa tiene una señal de baja frecuencia y una señal de alta frecuencia. Ambas señales son procesadas por una arquitectura de red neuronal y se recomponen posteriormente para generar pronósticos.

En este mismo sentido, Pryima, Vovk y Vovk (2019) utilizaron una arquitectura de perceptrón multicapa y otra con funciones de base radial (RBF Radial Basis Functions) para pronosticar los principales índices (UX Ukrainian Exchange y PFTS First Stock Trade System) del mercado de valores de Ucrania y demostraron cómo la transformación de las series temporales de entrada afecta significativamente el rendimiento de un modelo de red neuronal en la predicción de estos índices bursátiles.

Arango y Velásquez (2014) utilizaron una red neuronal autorregresiva para modelar la dinámica del Índice General de la Bolsa de Valores de Colombia (IGBC) y compararon sus resultados con modelos neurodifusos lineales y adaptativos. Sus principales conclusiones son que su propuesta permite manejar variables explicativas que contienen propiedades complejas como la heterocedasticidad y la no normalidad de los residuos, pero aún más importante, que es necesario considerar como entradas no solo las variables explicativas, sino también sus interacciones.

Para el caso del mercado de valores de Malasia, Sagir y Sathasivan (2017) proponen un modelo de predicción basado en redes neuronales artificiales no lineales (NARX Nonlinear Autoregressive with eXogenous input) para estimar los valores futuros de cinco índices diferentes y concluyen que sus predicciones son más precisas que un modelo basado en regresiones lineales múltiples. Esta misma arquitectura de redes NARX es utilizada por Dávila *et al.* (2017) para estimar el índice de la bolsa de valores de Ecuador (ECUINDEX), obteniendo un error promedio de menos de 0.25 % y un porcentaje de éxito en la predicción de la dirección de cambio mayor de 68 %.



## 2.4. Pronósticos en Mercados Financieros con Aprendizaje Profundo

Las técnicas convencionales de aprendizaje automático tienen una capacidad limitada para procesar datos en su forma original. Los modelos tradicionales requieren una ingeniería cuidadosa y una considerable experiencia en el dominio del problema para extraer las características y transformarlas en una representación interna adecuada o un vector de características desde el cual se pueda detectar o clasificar patrones. Para superar este inconveniente, en años recientes se han propuesto arquitecturas de redes neuronales denominadas de aprendizaje profundo.

Los métodos de aprendizaje profundo son técnicas de aprendizaje automático con múltiples niveles jerárquicos de representación, obtenidos mediante la composición de módulos simples, pero no lineales, que transforman los datos en un nivel de representación cada vez más abstracto, bajo el supuesto de que mediante la composición de suficientes transformaciones es posible aprender funciones aún más complejas (Lecun, Bengio, y Hinton, 2015).

Para los modelos de aprendizaje profundo, las capas superiores amplifican aspectos de la entrada que son importantes para la discriminación y suprimen variaciones irrelevantes. El aspecto clave del aprendizaje profundo es que estas capas de extracción de características no están diseñadas por humanos, sino que se aprenden de las relaciones entre los mismos datos a través de un procedimiento de aprendizaje automático de propósito general en los primeros niveles de la red.

Siguiendo esta idea, las arquitecturas de aprendizaje profundo han mejorado drásticamente en tareas como reconocimiento del habla, reconocimiento y clasificación de imágenes, detección de objetos o descubrimiento de fármacos, pero su aplicación en el campo de los pronósticos financieros aún no se encuentra muy extendida. Esto se explica debido a que los problemas en los mercados financieros son sumamente diferentes de las aplicaciones típicas de aprendizaje profundo en varios aspectos; uno de los más notables es que el énfasis no está en replicar tareas que los humanos ya hacen bien.

Sin embargo, las herramientas de aprendizaje profundo pueden ser útiles en esta área, ya que representan la mejor manera disponible de mapear funciones de entrada como el rendimiento de otros activos, datos macroeconómicos, información demográfica, etcétera, a resultados del mercado bursátil (Heaton, Polson, y Witte, 2016). Al menos en teoría, una red de aprendizaje profundo puede encontrar, sin importar cuán complejas y no lineales sean, las relaciones necesarias para predecir la trayectoria futura de un activo financiero.

Para el caso del índice S&P 500, X. Li *et al.* (2019) proponen una arquitectura de aprendizaje profundo denominada “Long Short Term Memory” (LSTM) que le permite incorporar información cuantitativa y análisis de sentimientos de noticias financieras. Los resultados demuestran que su modelo puede predecir el precio de las acciones con precisión y reflejar de manera acertada la tendencia general del mercado. Otras investigaciones, utilizan la arquitectura LSTM para aprender y pronosticar índices de valoración en mercados financieros (G. Li, Xiao, y Guo, 2019) y evalúan el sesgo de sus predicciones mediante la precisión de la tendencia pronosticada, la tasa de desviación de pronóstico promedio y el error cuadrático medio. Los resultados muestran que el modelo LSTM tiene un buen efecto predictivo sobre este tipo de secuencia de datos, confirmando el valor de las investigaciones teóricas y también su importancia como referencia para inversores del mercado bursátil.

Otra de las arquitecturas principales del aprendizaje profundo son las redes convolucionales (CNN Convolutional Neural Network). Cao y Wang (2019) las utilizaron para pronosticar el índice Hong Kong Hang Seng Index (HSI). Ellos proponen un modelo de predicción híbrido basado en CNN y máquinas de soporte vectorial (SVM Support-Vector Machines) para mejorar sus predicciones. Concluyen que la combinación de ambos algoritmos mejora significativamente la precisión de sus pronósticos y que este esquema puede ser útil incluso para realizar operaciones en tiempo real.

Niu *et al.* (2020) proponen un marco de trabajo genérico con el objetivo de capturar eficientemente la no linealidad inherente de las series de tiempo financieras. Su arquitectura consta de dos etapas: un modelo con tres unidades de aprendizaje profundo para selección de características y un modelo de corrección de errores. Para validar el desem-

peño de su arquitectura llevan a cabo casos de estudio de diferentes series financieras que incluyen el Dow Jones Industrial Average (DJIA), demostrando la superioridad de su arquitectura, en comparación con otros 16 modelos de referencia.

## 2.5. Pronósticos de Mercados Financieros en México con Aprendizaje Automático

En el caso del mercado financiero mexicano existen, comparado con otros mercados financieros emergentes, pocos trabajos sobre el tema.

Utilizando modelos econométricos tradicionales, López Herrera (2004) hace un análisis del comportamiento del IPC de la BMV mediante un modelo autorregresivo. Utiliza una ventana de doce años para su análisis y encuentra fuerte evidencia de que la volatilidad del IPC muestra un alto grado de persistencia en el tiempo. Para manejar la volatilidad, propone un modelo simétrico GARCH y dos modelos asimétricos TARCH y EGARCH, que incluyen un modelo autorregresivo. Concluye que el modelo EGARCH presenta mejores cualidades para pronosticar el IPC, comparado con los otros dos modelos, pero en general se encuentra evidencia de que los modelos, al controlar la volatilidad cambiante en el tiempo, reproducen adecuadamente las principales características de la distribución empírica del IPC para el horizonte del pronóstico.

Para manejar esta volatilidad, Cabrera y Ortiz (2012) proponen utilizar redes neuronales diferenciales (RND), una arquitectura empleada frecuentemente en análisis de procesos de biotecnología. Mediante esta técnica, estiman la evolución del rendimiento intradía del IPC en una ventana de tiempo de 20 años. Basados en su desempeño concluyen que el uso de estas redes es una opción viable y robusta para la descripción precisa de series de tiempo financieras. Adicionalmente, afirman que las redes neuronales diferenciales como sistema de predicción intradía, permiten obtener valores muy cercanos a la realidad, lo que brinda enormes posibilidades de generar beneficios económicos en la compra-venta diaria de activos bursátiles.

Un trabajo que combina el enfoque econométrico tradicional y la inteligencia arti-

ficial es el propuesto por León *et al.* (2018). Su objetivo es descomponer los factores de comportamiento del IPC y generar pronósticos mediante modelos econométricos y redes neuronales artificiales evolutivas. Reducen la complejidad de análisis y eliminan el ruido en la serie de datos del IPC mediante la descomposición empírica en modos (DEM) y el algoritmo de selección de características de programación evolutiva de redes (FS-EPNet). Los resultados experimentales permiten concluir que las predicciones generadas con esta metodología son más precisas que aquellas obtenidas a partir de los datos originales y confirma la necesidad de abordar el análisis del mercado bursátil con herramientas no paramétricas y no lineales.

# Capítulo 3

## Marco Teórico

### 3.1. Inteligencia Artificial

Han pasado ya casi 70 años desde que el matemático Alan Turing marcó el rumbo de la Inteligencia Artificial (IA) con una simple pregunta: “¿Pueden pensar las máquinas?” (Turing, 1950). Su trabajo estableció el objetivo fundamental y la visión general de la IA, al intentar replicar o simular la inteligencia humana en máquinas. Esto llevó al surgimiento de un nuevo campo de investigación que se fundó como disciplina académica en 1956, en una conferencia en el Dartmouth College, en New Hampshire, donde el término IA fue acuñado por John McCarthy.

La definición de IA es amplia, ya que involucra disciplinas como las matemáticas, la ingeniería, la neurociencia o la filosofía, pero podemos sintetizar su objetivo como el estudio del comportamiento inteligente en las máquinas. Este comportamiento inteligente incluye percibir, razonar, comunicarse y tomar decisiones en medio de ambientes complejos (Russel y Norvig, 2010). Coloquialmente, el término IA se aplica cuando una máquina imita las funciones cognitivas que los humanos vinculan con la racionalidad, por ejemplo, la resolución de problemas.

Hoy en día, el término IA se asigna a algoritmos que responden a estímulos externos con acciones asociadas a la mente humana, como juicio e intención. Estos sistemas de software “toman decisiones que normalmente requieren un nivel humano de experiencia, ayudando a las personas a anticipar problemas o lidiar con ellos a medida que surgen”

(Shubhendu y Vijay, 2013). Además, estos sistemas cuentan con tres cualidades que constituyen la esencia de la inteligencia artificial: intencionalidad, inteligencia y adaptabilidad.

Durante su desarrollo, la IA ha sido abordada desde diferentes aproximaciones, pero los principales paradigmas pueden clasificarse en dos grupos. El primer grupo se denomina “Inteligencia Artificial Simbólica” e incluye aproximaciones basadas en procesamiento de símbolos. Este enfoque representa el conocimiento sobre un dominio con sentencias declarativas, basadas en la lógica de predicados (Newell, 1982) y obtiene sus conclusiones mediante mecanismos de inferencia similares a los que utilizan los seres humanos para razonar.

El segundo grupo se denomina “Inteligencia Artificial Subsimbólica” y sigue un estilo de diseño ascendente, comenzando su implementación en los niveles más bajos y procediendo hacia los niveles superiores. De acuerdo a este enfoque, primero es necesario desarrollar las capacidades de procesamiento de señales y control presentes en la mente humana, lo que permitirá crear un sustrato sobre el cual pueden construirse los niveles superiores de la inteligencia. Esta aproximación no intenta crear programas como simples autómatas que obedecen reglas concretas, sino más bien, reproducir los mecanismos básicos que nos permiten actuar como seres inteligentes, independientemente del contexto o del problema específico. Uno de los ejemplos más claros de este paradigma son las redes de neuronas artificiales (ANN Artificial Neural Networks).

## **3.2. Redes Neuronales Artificiales**

Las ANN no son un concepto nuevo. Su desarrollo más básico se remonta a mediados del siglo pasado (McCulloch y Pitts, 1943) y su formalización aparece en el trabajo de Rosenblatt (1958) con la definición del algoritmo del perceptrón.

### **3.2.1. El Perceptrón**

El perceptrón es el bloque fundamental sobre el cual se construyen las ANN y su funcionamiento es relativamente sencillo. La información se propaga siempre hacia adelante

(por lo que generalmente se denominan redes *feed-forward*) iniciando desde las neuronas de entrada que reciben un vector de dimensión  $m$  ( $x_1, x_2, \dots, x_m$ ). A su vez, cada una de estas entradas tiene asociado un peso sináptico  $w_1, w_2, \dots, w_m$  que multiplica al vector de entradas correspondiente. Los resultados se suman y se hacen pasar a través de una función generalmente no lineal, dando lugar al resultado final  $\hat{y}$ .

En la práctica, generalmente se agrega una neurona adicional con un valor constante para agregar un sesgo. El objetivo de este sesgo es permitir desplazar la función, independientemente de las entradas y los pesos sinápticos (Haykin, 1999).

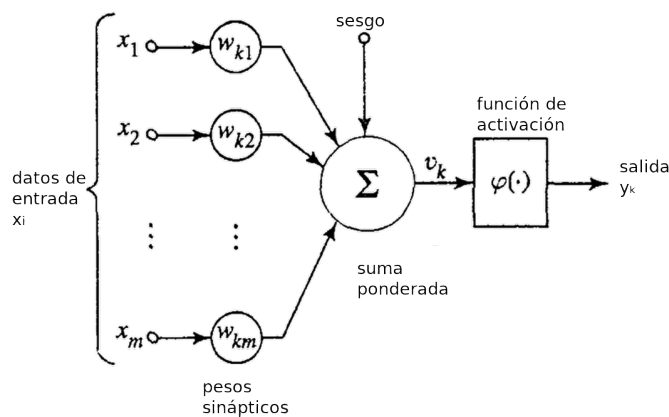


Figura 3.1: Perceptrón

Fuente: Haykin (1999)

En términos matemáticos, el perceptrón es un clasificador binario basado en una función de umbral  $f(x)$  que mapea un vector de entradas  $x$  a un valor binario, es decir:

$$f(x) = \begin{cases} 1 & \text{si } w \cdot x + w_0 > 0 \\ 0 & \text{si } w \cdot x + w_0 \leq 0 \end{cases} \quad (3.1)$$

Donde  $w$  es un vector de pesos y  $w \cdot x$  es la operación producto punto entre vectores. En esta definición,  $w_0$  es el sesgo que desplaza la frontera de decisión y no depende de ningún valor de entrada. Esta misma idea también puede expresarse de forma más general como:

$$f(x) = \mathfrak{F} \left( w_0 + \sum_{i=1}^n w_i \cdot x_i \right) \quad (3.2)$$

Donde  $\mathfrak{F}$  es una función de tipo escalón unitario.

Aunque el perceptrón inicialmente parecía prometedor, se demostró que no podía clasificar patrones no lineales (Minsky y Papert, 1969), provocando que el campo de la investigación de las ANN se estancara durante las siguientes décadas.

### 3.2.2. Funciones de Activación

La función de activación es el componente que determina si una neurona debe activarse o no. El propósito más importante de la función de activación es introducir el efecto de no linealidad en la salida de una neurona, pues, como se planteó previamente, una red neuronal sin una función de activación es solo un modelo de regresión lineal.

La introducción del efecto de no linealidad es fundamental ya que, en la mayoría de los problemas prácticos del mundo real, los datos de entrada son no lineales.

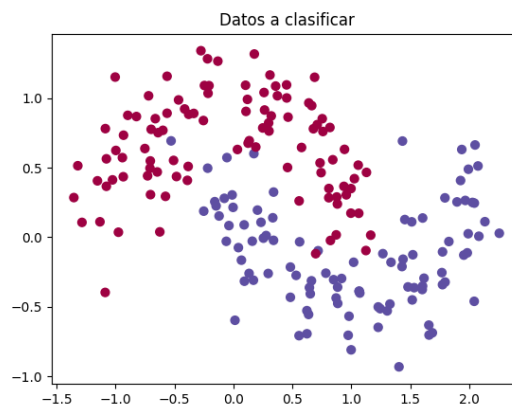


Figura 3.2: Datos no linealmente separables

Fuente: Britz (2015)

Para datos como los mostrados en la Figura 3.2, sin importar cuánto esfuerzo computacional se destine, es imposible obtener una clasificación perfecta si se utiliza una función lineal. En estos casos, la mejor clasificación siempre se obtendrá mediante funciones complejas no lineales 3.3.

Por otra parte, la función de activación es un componente crucial en la arquitectura de la ANN, que determina el resultado de un modelo, su precisión y también su eficiencia computacional por el importante efecto que tiene en la capacidad y velocidad de convergencia en la etapa de entrenamiento de la red.



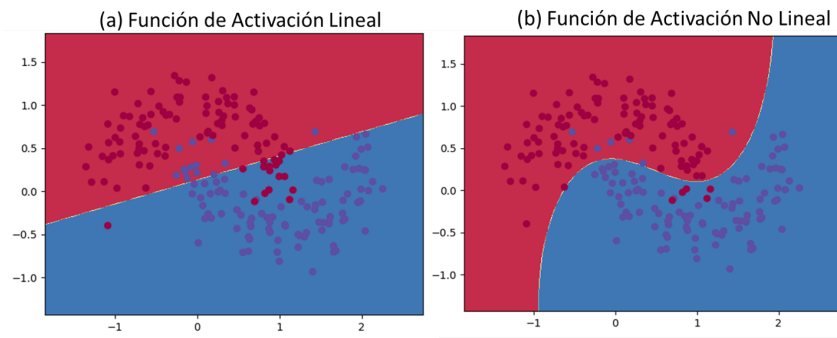


Figura 3.3: Clasificación con funciones de activación lineal y no lineal

Fuente: Britz (2015)

Aunque dentro del aprendizaje automático existe una gran cantidad de funciones de activación, en la práctica las más utilizadas por sus características no lineales son la función sigmoïdal, la tangente hiperbólica y la función ReLU.

### 3.2.3. Perceptrón Multicapa

Las ANN tuvieron un fuerte resurgimiento como herramientas de reconocimiento de patrones después de que se demostrara que arquitecturas neuronales apiladas pueden aprender mapas funcionales complejos no lineales. Simultáneamente, Cybenko (1989) y Hornik (1989) demostraron que un perceptrón multicapa (MLP Multilayer Perceptron), con al menos una capa oculta, es un aproximador universal de cualquier función continua sobre  $\mathbb{R}^n$ , por lo que agregar una segunda capa de perceptrones es suficiente para resolver muchos problemas que de otro modo no serían separables: una red neuronal multicapa puede distorsionar el espacio de entrada para hacer que las clases de datos sean linealmente separables y clasificarlos exitosamente.

La manera en que las células se conectan entre sí es denominada patrón de conectividad o arquitectura de la red. La estructura básica de interconexión es la de red multicapa completamente conectada (*fully connected*), por lo que también reciben el nombre de redes neuronales densamente conectadas o simplemente redes neuronales densas (Fig. 3.4). El primer nivel lo constituyen las células de entrada; estas unidades no actúan necesariamente como perceptrones, sino que reciben los valores mediante vectores y se encargan de iniciar la propagación a través de toda la red. A continuación, hay una o varias capas

intermedias, llamadas *capas ocultas*, cuyas unidades procesan patrones particulares que aparecen en los datos de la capa de entrada. El último nivel sirve como salida de toda la red.

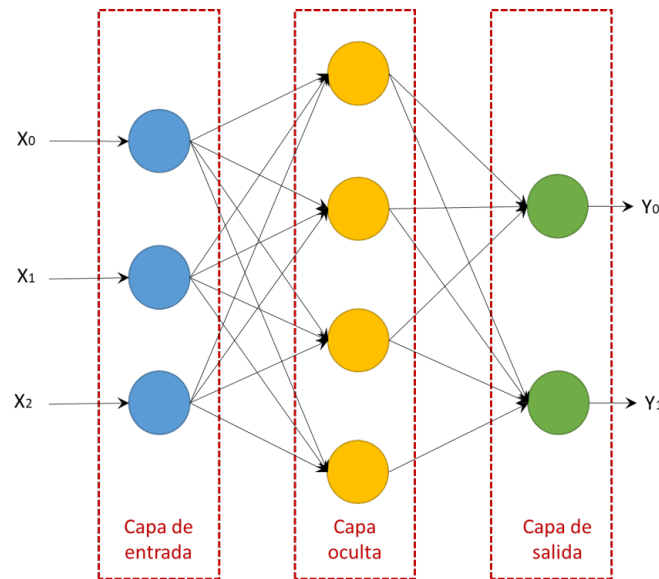


Figura 3.4: Red neuronal densa

Fuente: Elaboración propia.

El funcionamiento de una red de neuronas densamente conectada puede describirse mediante la ecuación:

$$\bar{Y} = \mathfrak{F}(\mathfrak{F}(\bar{X} \cdot \bar{W}_1) \cdot \bar{W}_2) \quad (3.3)$$

Donde  $\bar{X}$  es el vector de entradas,  $\bar{W}_1$  y  $\bar{W}_2$  son los vectores de pesos de la primera y segunda capa,  $\mathfrak{F}$  es la función de activación de las neuronas y  $\bar{Y}$  es el vector de salidas que produce la red. Los MLP también pueden definirse como grafos cuyos nodos están formados por unidades de procesamiento idénticas que propagan la información a través de todos los arcos.

Sin embargo, aunque el MLP sea una de las arquitecturas más conocidas y utilizadas, esto no significa que sea la más potente o con mejores resultados en todas las áreas de aplicación, ya que tienen una función de pérdida en general no convexa con más de un mínimo local, por lo que los valores aleatorios inicialmente asignados a los pesos de los enlaces entre neuronas pueden conducir a modelos con diferente precisión (Hinton, 1989).

### 3.2.4. Algoritmo de Aprendizaje

La característica clave de las redes neuronales artificiales es su “capacidad de aprender sin ser programadas explícitamente” (Awad y Khanna, 2015). Si ya se conoce un modelo matemático conveniente que describe el conjunto de datos, es poco probable que se necesite una red neuronal, pero cuando las reglas que subyacen a los datos se conocen solo parcialmente, o definitivamente no se conocen, una red neuronal permite descubrir relaciones entre los datos y reflejar interrelaciones complejas.

Cuando la red neuronal se encuentra en un contexto de aprendizaje supervisado, es entrenada con pares de entrada-salida y se busca que aprenda la relación o dependencia que existe entre ellas. En el entrenamiento se ajustan los valores de los pesos sinápticos con el objetivo de minimizar una función de pérdida empírica que se define como:

$$\mathcal{J}(W) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f(x^{(i)}; W), y^{(i)}) \quad (3.4)$$

Donde  $n$  es el total de observaciones en el conjunto de entrenamiento,  $y^{(i)}$  es el valor real asociado a las observaciones y  $f(x^{(i)}; W)$  es el valor estimado por la red mediante el vector de entradas  $x^{(i)}$  y los pesos sinápticos  $W$  de toda la red neuronal.

Por lo tanto, el problema de aprendizaje en las redes neuronales se formula en términos de la minimización de la función pérdida asociada, que puede escribirse simplemente como  $\mathcal{J}(W)$  porque el error que comete la red neuronal depende únicamente de los pesos sinápticos asociados a la misma. El objetivo de esta formalización es encontrar entonces el valor  $W^*$  para el cual se obtiene un mínimo global de la función  $\mathcal{L}$ , convirtiendo el problema de aprendizaje en un problema de optimización.

$$W^* = \arg \min_W \mathcal{J}(W) \quad (3.5)$$

En general, la función de pérdida es una función no lineal, por lo que no se dispone de algoritmos sencillos y exactos para encontrar sus mínimos. En consecuencia, se hace uso de métodos heurísticos de búsqueda a través del espacio de parámetros que, idealmente, se aproximen de forma iterada a un error mínimo de la red para los parámetros adecuados.

En la práctica, la heurística utilizada es el *método del gradiente descendente* y el algoritmo que permite comunicar los errores medidos en la salida hacia las neuronas intermedias para ajustar los pesos sinápticos de toda la red se conoce como *algoritmo de retro-propagación* (Rumelhart, Hinton, y Williams, 1986).

### 3.3. Redes Neuronales Profundas

El rendimiento de las redes neuronales (y en general de los métodos de aprendizaje automático) dependen en buena medida de elegir una forma de representación adecuada para los datos de entrada, según el campo específico del problema que se aborda. Por esa razón, gran parte del esfuerzo en la implementación de algoritmos de aprendizaje automático se dedica al preprocesamiento y transformaciones de datos que dan como resultado una representación de los datos que permiten un aprendizaje efectivo. Dicha ingeniería de características es fundamental, pero requiere mucho trabajo ya que los algoritmos de aprendizaje actuales son incapaces de extraer y organizar la información discriminatoria de los datos en su forma cruda (Bengio, Courville, y Vincent, 2013).

Una IA debe comprender fundamentalmente el conocimiento del contexto en el que se encuentra y esto solo se puede lograr si es capaz de aprender a identificar los factores explicativos subyacentes que se encuentran ocultos en los datos de entrada. Sin embargo, gran parte de estos factores de conocimiento son subjetivos o intuitivos y, por lo tanto, difíciles de estructurar de manera formal en términos de una representación computacional. Este es actualmente uno de los desafíos clave en la IA.

El principal enfoque del aprendizaje profundo es permitir que las computadoras aprendan este conocimiento a través de la experiencia (datos de entrenamiento) y comprendan el problema en términos de descomponer conceptos complejos en una jerarquía hasta llegar a representar conceptos más simples y las relaciones que existen entre ellos. Esta jerarquía permite al algoritmo aprender conceptos complicados al construirlos a partir de otros más básicos. En términos de un grafo puede decirse que las representaciones se van construyendo una encima de otra, dando lugar a un grafo profundo compuesto de muchas capas. El aprendizaje de la representación de los datos lo realiza la arquitectura

del modelo.

Por esta razón se denomina a este enfoque “aprendizaje profundo” y se considera un área dentro de la representación del conocimiento.

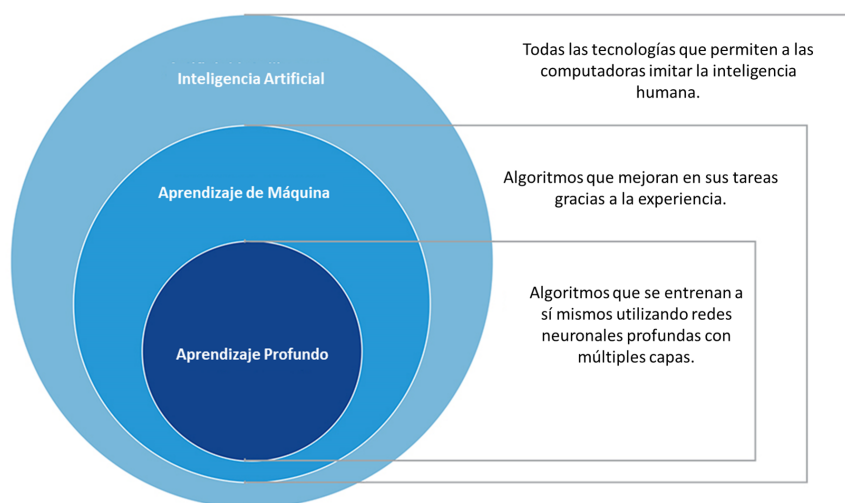


Figura 3.5: Ubicación del aprendizaje profundo

Fuente: Elaboración propia.

En su forma más simple, el aprendizaje profundo es una implementación de redes neuronales con varias capas de nodos (4 o más) entre las capas de entrada y salida (Arel, Rose, y Karnowski, 2010). La representación del conocimiento se forma mediante la composición de múltiples transformaciones no lineales de los datos de entrada con el objetivo de producir representaciones abstractas y útiles para las tareas de clasificación o predicción.

Por ejemplo, en muchos modelos de aprendizaje profundo, hay una capa lineal simple apilada encima de un bloque complejo con muchas capas no lineales de diferentes tipos. En general, se debe a que el bloque complejo transforma la entrada en una representación que solo requiere una capa lineal simple para realizar su tarea específica. Sin la transformación realizada por el bloque complejo, no sería posible extraer las características abstractas clave y útiles para el aprendizaje.

En la última década, el aprendizaje profundo se ha aplicado con éxito a diversas tecnologías y mecanismos que requieren grandes volúmenes de datos para la etapa de entrenamiento (Chen y Lin, 2014) en una enorme variedad de dominios. Antes de 2006, aunque ya existían las bases del aprendizaje profundo, los investigadores eran escépticos

sobre su viabilidad y éxito Bengio *et al.*(2007). Actualmente, muchas tecnologías utilizan las técnicas de aprendizaje profundo para lograr extraer información útil de los datos y hacer que una máquina sea autosuficiente para imitar algunas tareas humanas con un desempeño muy superior, como sucede hoy en día en la clasificación de imágenes.

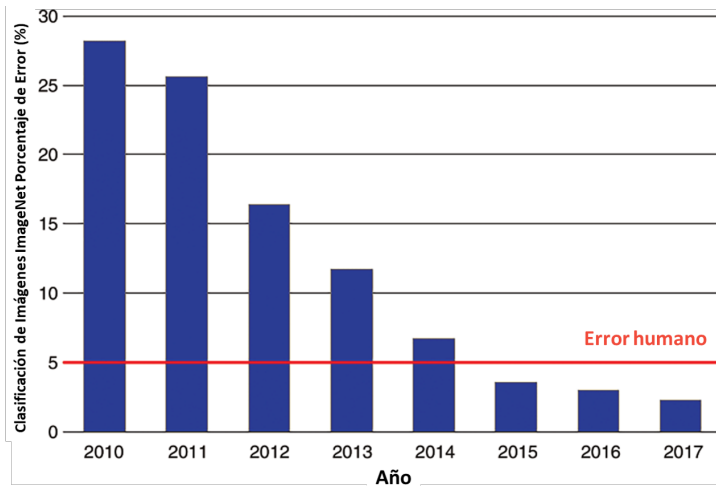


Figura 3.6: Error en clasificación de imágenes

Fuente: Elaboración propia con datos de ImageNet.

Actualmente, la humanidad está generando y acumulando un gran depósito de datos que pronto superará nuestra capacidad computacional (Gantz y Reinsel, 2010). En tal escenario, el aprendizaje profundo se proyecta como la técnica de IA capaz de manejar esta vasta información y ponerla en uso de manera efectiva. A pesar de su aplicación generalizada, el aprendizaje profundo todavía está en sus etapas iniciales y se ha convertido rápidamente en un tema central de investigación para desarrollar máquinas inteligentes y autónomas.

### 3.3.1. Redes Neuronales Recurrentes

Uno de los principales problemas a los que se enfrentan las redes neuronales básicas es la persistencia de la información cuando se analizan datos secuenciales.

En un principio, podría plantearse utilizar un perceptrón regular y repetirlo en varias ocasiones, pero el orden en el que aparecen los datos secuenciales en la entrada significa algo: un elemento de entrada en la serie está relacionado con otros en forma de una secuencia y muy probablemente influye en sus vecinos. De lo contrario, son solo “muchas”

entradas y no necesariamente una entrada en “serie” (Lipton, Berkowitz, y Elkan, 2015).

Por lo tanto, se requiere una arquitectura que capture esta relación entre las entradas de manera significativa. Las redes neuronales recurrentes o RNN (Recurrent Neural Network), son un tipo de redes neuronales cuya principal característica es su capacidad de tratar con datos en forma de serie o secuencia.

Si bien las RNN aprenden de manera similar al perceptrón básico durante el entrenamiento, tienen la característica de recordar lo aprendido de entradas anteriores al generar salidas. Las RNN pueden tomar uno o más vectores de entrada y producir uno o más vectores de salida y las salidas están influenciadas no solo por los pesos aplicados a las entradas, sino también por un vector de estado “oculto” que representa el contexto basado en la secuencia que tienen entradas y salidas. De esta forma, la misma entrada podría producir una salida diferente dependiendo de las entradas anteriores en la secuencia.

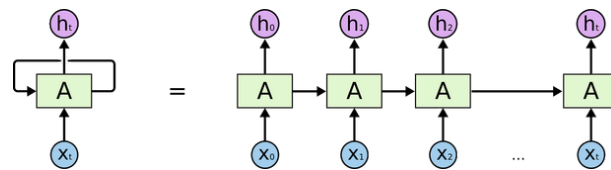


Figura 3.7: Red neuronal recurrente

Fuente: Olah (2015)

En una RNN la idea fundamental es que la información o características extraídas por la red en el instante de tiempo  $t$  afecta a la información o características que se aprenden en los instantes  $t, t+1, t+2, \dots$  y son almacenados en el estado  $h$  de la unidad recurrente. En la Figura 3.7,  $x_{(t)}$  es la entrada en el paso de tiempo  $t$ ,  $h_{(t)}$  es el estado oculto en el paso de tiempo  $t$  (la “memoria” de la red) y por lo tanto, se calcula en función del estado oculto anterior y la entrada en el paso actual:

$$h_t = f(U \cdot x_t + W \cdot h_{t-1}) \quad (3.6)$$

La función  $f$  generalmente no es lineal, como  $\tanh$  o  $ReLU$ . En la práctica, el primer estado oculto que se requiere para arrancar los cálculos generalmente se inicializa en cero. Finalmente,  $o_t$  es la salida en el paso de tiempo  $t$  y se calcula como:

$$o^t = \sigma(W \cdot h_t) \quad (3.7)$$

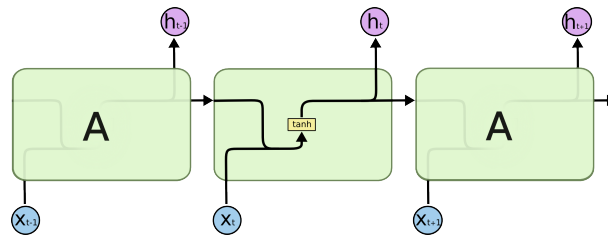


Figura 3.8: Red neuronal recurrente con función de activación  $\tanh$

Fuente: Olah (2015)

En ocasiones, solo se necesita la información reciente de una secuencia para pronosticar el valor actual. En estos casos donde la brecha entre la información relevante y el momento del pronóstico es pequeña, las RNN básicas pueden funcionar bastante bien (Figura 3.9).

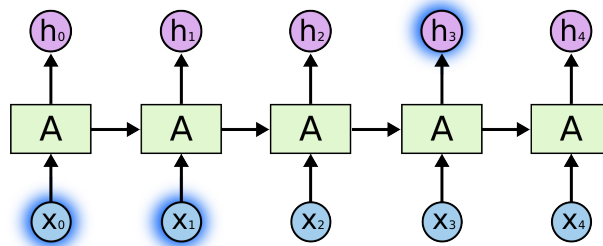


Figura 3.9: RNN con dependencias de corto plazo

Fuente: Olah (2015)

Sin embargo, existen casos en los que se necesita un contexto para realizar una predicción. A medida que crece esa brecha, las RNN básicas no pueden transmitir este conocimiento a las unidades posteriores para conectar este conocimiento (Figura 3.10), un problema asociado principalmente a un fenómeno denominado *desvanecimiento del gradiente* (Bengio, Simard, y Frasconi, 2014).

### 3.3.2. Redes de Memoria a Corto y Largo Plazo (LSTM)

Las redes LSTM (Long and Short Term Memory) son un tipo especial de RNN capaces de aprender dependencias a largo plazo. Fueron introducidos por Hochreiter y Schmidhuber



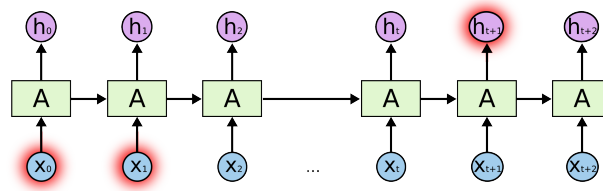


Figura 3.10: RNN con dependencias de largo plazo

Fuente: Olah (2015)

(1997) y demostraron un buen funcionamiento en una gran variedad de problemas con datos secuenciales y en la actualidad son ampliamente utilizadas.

Las LSTM están diseñadas específicamente para solucionar el problema de dependencia a largo plazo, por lo que recordar información durante largos períodos de tiempo es su comportamiento predeterminado. Todas las redes neuronales recurrentes tienen la forma de una cadena de módulos repetitivos de red neuronal. En los RNN estándar, este módulo repetitivo tiene una estructura simple con una sola capa *tanh*.

Las redes LSTM también tienen esta estructura tipo cadena (Figura 3.11), pero el módulo de repetición tiene una estructura diferente: en lugar de tener una sola capa de red neuronal, hay cuatro que interactúan de una manera particular.

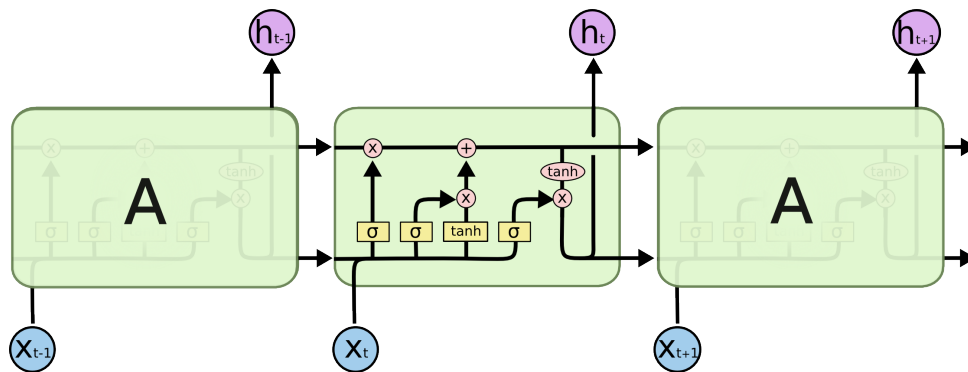


Figura 3.11: Secuencia de unidades LSTM

Fuente: Olah (2015)

La unidad LSTM tiene la capacidad de eliminar o agregar información al estado de la célula, regulado por estructuras llamadas puertas. Las puertas son un mecanismo para dejar pasar la información y se componen de una capa de red neuronal *sigmoidal* y una operación de multiplicación elemento a elemento.

- Capa de puerta de olvido

El primer paso en la unidad LSTM es decidir qué información se va a olvidar de estado de la celda. Esta decisión la toma una capa *sigmoidal* llamada “capa de la puerta de olvido” (Figura 3.12). Toma los datos  $h_{t-1}$  y  $x_t$ , y genera un número entre 0 y 1 para cada valor en el estado de la celda  $C_{t-1}$ . Un 1 representa mantener completamente la información, mientras que un 0 representa olvidar completamente la información.

$$f_{(t)} = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.8)$$

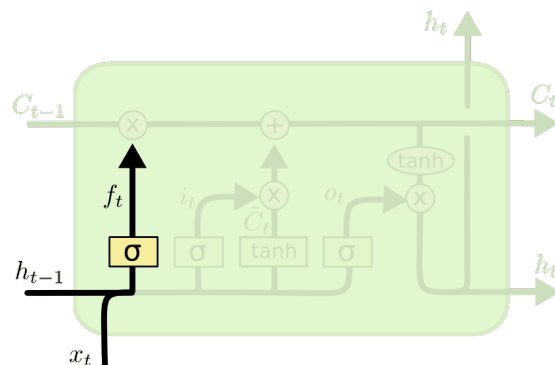


Figura 3.12: Capa de la puerta de olvido

Fuente: Olah (2015)

- Capa de puerta de entrada

El siguiente paso es decidir qué nueva información se va a almacenar en el estado de la celda actual. Esto tiene dos etapas: primero, una capa *sigmoidal* llamada “capa de puerta de entrada” (Figura 3.13) decide qué valores se actualizarán. Luego, una capa *tanh* crea un vector de nuevos valores candidatos  $\tilde{C}_t$  que podrían agregarse al estado. En el siguiente paso, se combinan para crear la actualización del estado.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.9)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.10)$$

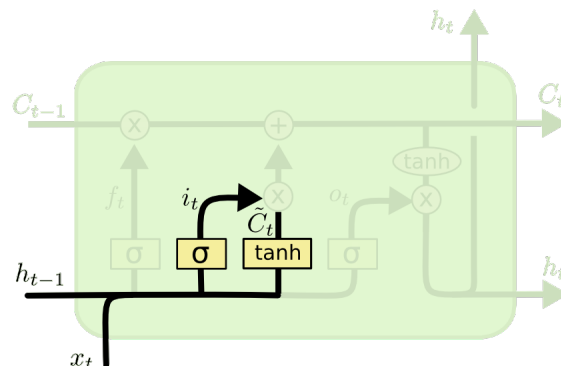


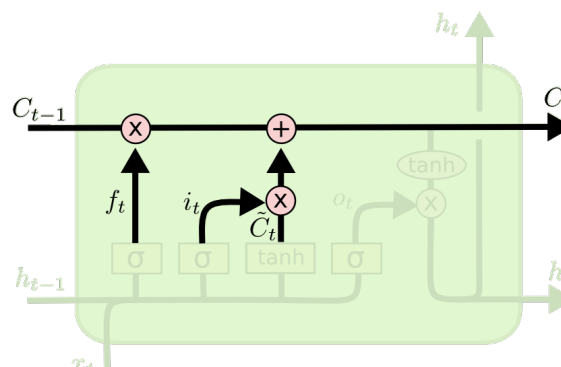
Figura 3.13: Capa de la puerta de entrada

Fuente: Olah (2015)

- Capa de puerta de actualización

Después se actualiza el antiguo estado de la celda  $C_{t-1}$  en el nuevo estado de la celda  $C_t$  con base en los datos calculados previamente. Se multiplica el estado anterior  $C_{t-1}$  por  $f_t$ , olvidando las cosas que se deben olvidar y luego se agrega  $i_t * \tilde{C}_t$  (Figura 3.14). Estos son los nuevos valores candidatos, escalados de acuerdo a cuánto se decidió actualizar.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3.11)$$

Figura 3.14: Actualización del estado  $C_t$ 

Fuente: Olah (2015)

- Capa de puerta de salida

Finalmente, se calcula el valor que la celda va a generar como salida (Figura 3.15). Esta salida se basará en el estado de celda, pero será una versión filtrada. Primero, se ejecuta una capa *sigmoidal* que decide qué partes del estado de la celda se va a generar.

Luego, se pasa el estado de la celda a través de una capa *tanh* para escalar los valores entre -1 y 1 y se multiplican por la salida de la capa *sigmoidal*.

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (3.12)$$

$$h_t = o_t * \tanh(C_t) \quad (3.13)$$

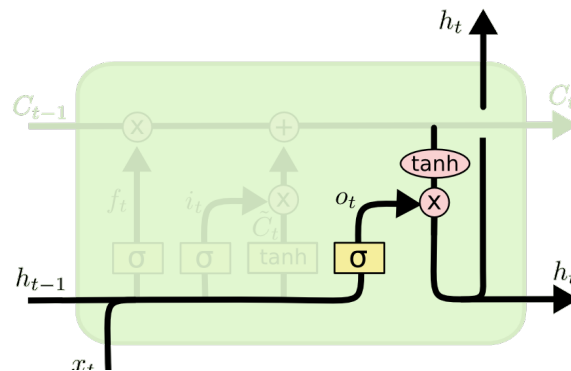


Figura 3.15: Salida de la unidad LSTM

Fuente: Olah (2015)

### 3.3.3. Redes Neuronales Convolucionales

Las redes neuronales convolucionales (CNN Convolutional Neural Network), son un tipo de red neuronal profunda especializada en procesar datos que tienen una topología similar a una matriz. Desde la década de 1980 existían ideas sobre la implementación de las CNN, pero es hasta finales del siglo XX que se presentan las ideas más importantes de esta tecnología con el trabajo seminal de Yann LeCun (1998). El nombre de red neuronal convolucional indica que la red emplea la operación matemática llamada convolución, que es un tipo especializado de operación lineal. Las redes convolucionales son redes neuronales que usan la convolución, en lugar de la multiplicación matricial general, en al menos una de sus capas.

Los datos de entrada pueden incluir series temporales, que pueden considerarse como una cuadrícula unidimensional con muestras a intervalos de tiempo regulares, datos de imágenes, que pueden considerarse como una cuadrícula bidimensional y datos en tres

dimensiones como pueden ser escáneres de resonancia magnética o videos (dos dimensiones asociadas a las imágenes más una dimensión asociada al desarrollo temporal del video).

En su forma más general, la convolución es una operación de dos funciones con un argumento con valor real y se define como:

$$s(t) = \int x(a)w(t-a)da \quad (3.14)$$

La operación de convolución se denota en su forma sencilla simplemente con un asterisco:

$$s(t) = (x * w)(t) \quad (3.15)$$

En la terminología de red convolucional, el primer argumento de la convolución  $x$  a menudo se denomina entrada y el segundo argumento (la función  $w$ ) se denomina como el núcleo (*kernel*). La salida recibe el nombre de mapa de características.

Por lo general, cuando los datos son representados en computadora, el tiempo es discretizado, por lo que el índice de tiempo  $t$  solo puede tomar valores enteros. Si ahora asumimos que  $x$  y  $w$  se definen solo para los valores enteros de  $t$ , podemos definir la convolución discreta como:

$$s(t) = \sum_{a=-\infty}^{\infty} (x * w)(t) \quad (3.16)$$

Finalmente, si se utilizan convoluciones sobre más de un eje a la vez como, por ejemplo, en una imagen bidimensional, se utiliza un núcleo bidimensional, al que generalmente se denota con  $K$ .

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n) \quad (3.17)$$

Dado que la operación de convolución cumple con la propiedad de conmutatividad, podemos escribir de manera equivalente

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (3.18)$$

Esta última fórmula es más sencilla de implementar en términos computacionales, porque hay menos variación en el rango de valores válidos de  $m$  y  $n$ . Sin embargo, muchas bibliotecas de redes neuronales implementan la función llamada correlación cruzada, que es equivalente a la convolución pero sin voltear el núcleo, es decir:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (3.19)$$

En la práctica, la convolución es el proceso de transformar una matriz aplicando un *kernel* sobre cada punto y sus vecinos locales. El *kernel* es una matriz cuyo tamaño y valores determinan el efecto de transformación del proceso de convolución, que va procesando ventanas de datos y da como resultado un mapa de características (Fig. 3.16).

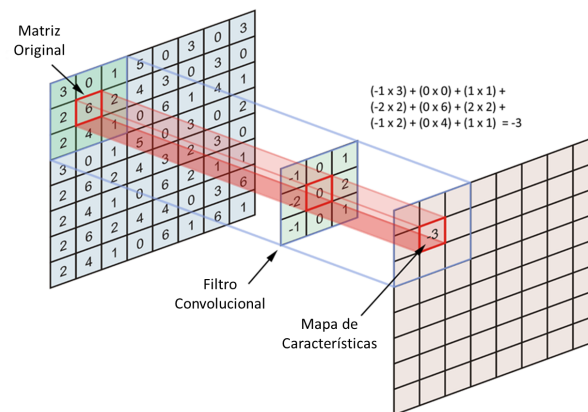


Figura 3.16: Ejemplo operación convolución en 2D

Fuente: Hartley (2014)

En las redes convolucionales, los parámetros de la capa de convolución consisten en un conjunto de *kernels* que se pueden aprender. Cada *kernel* es una matriz pequeña que se extiende por todos los datos de entrada y se calibra a través del entrenamiento. Por esta razón este tipo de redes tienen tanto éxito en la extracción de características en imágenes y clasificación de objetos.

Aplicando este mismo principio a convoluciones en una dimensión, puede observarse su equivalencia con la aplicación de filtros *kernel* para suavizamiento en series de tiempo.

# Capítulo 4

## Metodología

### 4.1. Datos

Con el objetivo de tener una mayor presencia internacional, la BMV firmó en mayo de 2015 un acuerdo con S&P Dow Jones Índices (S&P DJI) para el cálculo y operación de sus índices. Desde esa fecha, S&P DJI es el responsable de comercializar tanto los índices como la información al cierre del día de los mismos, mientras que la BMV comercializa la información de los índices en tiempo real. La finalidad es que los mercados mexicanos y sus participantes cuenten con más indicadores, diseñados de acuerdo a los estándares internacionales. Por esta razón, a partir del lunes 5 de junio de 2017, el IPC es también conocido bajo el nombre de S&P/IPC (BMV y S&P Dow Jones Índices, 2017).

Como parte de este acuerdo, también se presentaron cambios en la metodología del cálculo de los índices que incluyen el criterio de selección, el cambio en la muestra, el rebalanceo semestral, las fechas de referencia, reemplazos y acciones flotantes, entre otros. Esta nueva metodología fue aplicada a partir del lunes 18 de septiembre de 2017 al inicio de la sesión.

Para efectos del presente trabajo, los cambios más relevantes de la metodología son los relacionados con la selección de las acciones consideradas para el cálculo del índice, ya que ahora el universo elegible está compuesto por todas las series accionarias listadas en la BMV (excepto FIBRAS y Fideicomisos Hipotecarios), siempre y cuando cumplan con los criterios mínimos de Capitalización de Mercado y Liquidez (S&P Dow Jones Índices,

2014), hasta llegar a 35 series accionarias.

Si hay más de 35 series accionarias seleccionadas, las elegibles son ordenadas por su Valor de Capitalización de Mercado Flotante y la Mediana mensual de importe operado de 6 meses, las series accionarias con las mayores calificaciones son seleccionadas hasta que el índice alcance la muestra de 35 series. Si existe un empate en las calificaciones, la serie accionaria con la mayor liquidez con base en la Mediana mensual del importe operado de 6 meses es seleccionada.

El cambio de muestra del índice se realiza semestralmente en la apertura de mercado del lunes siguiente al tercer viernes de marzo y septiembre de cada año. Adicionalmente, es rebalanceado en la apertura de mercado del lunes siguiente al tercer viernes de junio y diciembre de cada año.

En este trabajo, la muestra considerada es la vigente desde marzo de 2019, presentada sin cambios para septiembre de 2019 y vigente hasta marzo de 2020. Las series accionarias presentes en esta muestra aparecen en el Apéndice 1 en orden alfabético, de acuerdo con su clave de cotización.

Todas las series utilizadas cuentan con datos del 2 de enero de 2017 al 31 de diciembre de 2019 (3 años), un total de 753 observaciones. Se considera el precio de cierre ajustado diario.

#### **4.1.1. Selección de Series Características**

En la Figura 4.1 se presenta el mapa de correlación del IPC y sus 35 componentes, empleando el coeficiente de correlación de Pearson. Los colores más intensos muestran aquellas series que presentan mayor correlación, tanto positiva como negativa, con el índice a estimar. También pueden observarse las fuertes interrelaciones que existen entre las mismas series accionarias. La información completa de los índices de correlación de la matriz se presenta en el Apéndice 2.



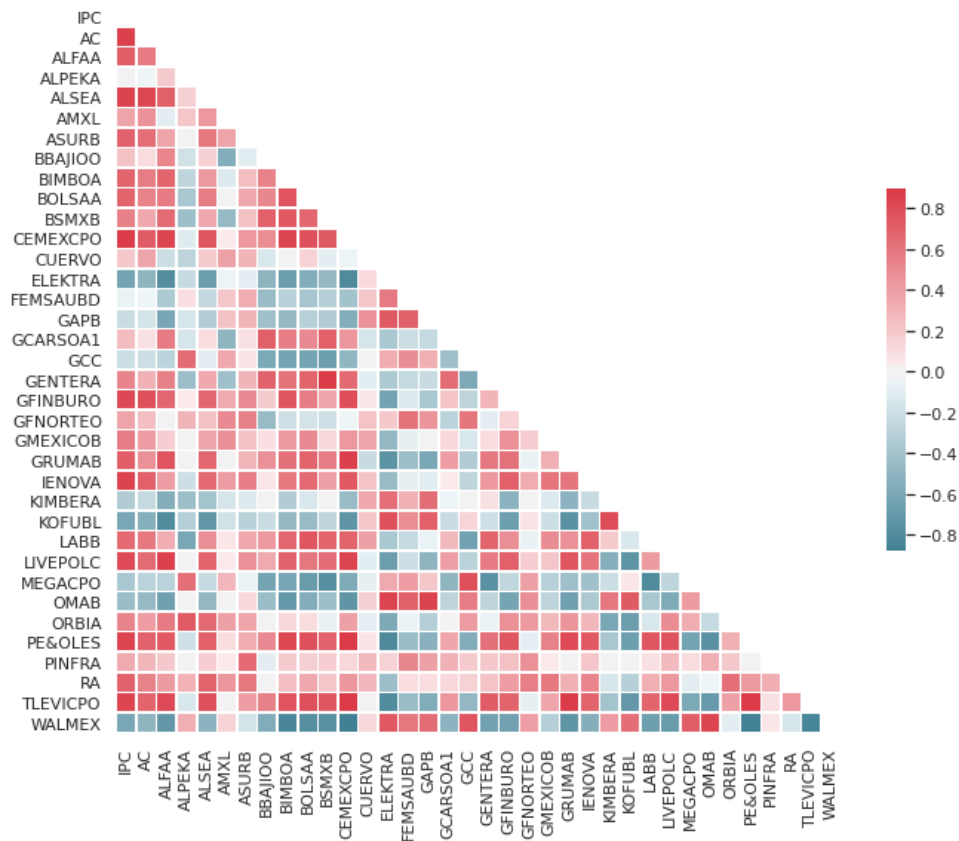


Figura 4.1: Correlación IPC y series accionarias

Fuente: Elaboración propia.

Un problema general de las técnicas multivariadas es conocido como “*la maldición de la dimensionalidad*”, el cual provoca que la eficiencia y la precisión del desempeño de los algoritmos de aprendizaje automático disminuya rápidamente a medida que aumenta el número de dimensiones o variables (Donoho, 2000). Este problema se ve incrementado en las últimas décadas con las innovaciones tecnológicas que han aumentado la capacidad para obtener datos de forma masiva.

Con el fin de reducir la dimensionalidad, se aplicó el criterio de correlación, distancia y asociación para determinar aquellas series que son más relevantes para el pronóstico, con base en pruebas estadísticas univariadas y fijando un umbral de correlación del 60%. Las series que califican para ser empleadas en este estudio según los criterios descritos son únicamente 21 y se muestran en la Figura 4.2. Adicionalmente, al analizar la calidad de los datos disponibles, se decidió omitir la serie accionaria *BSMXB.MX*, ya que dispone de información únicamente a partir del 27 de septiembre de 2017.

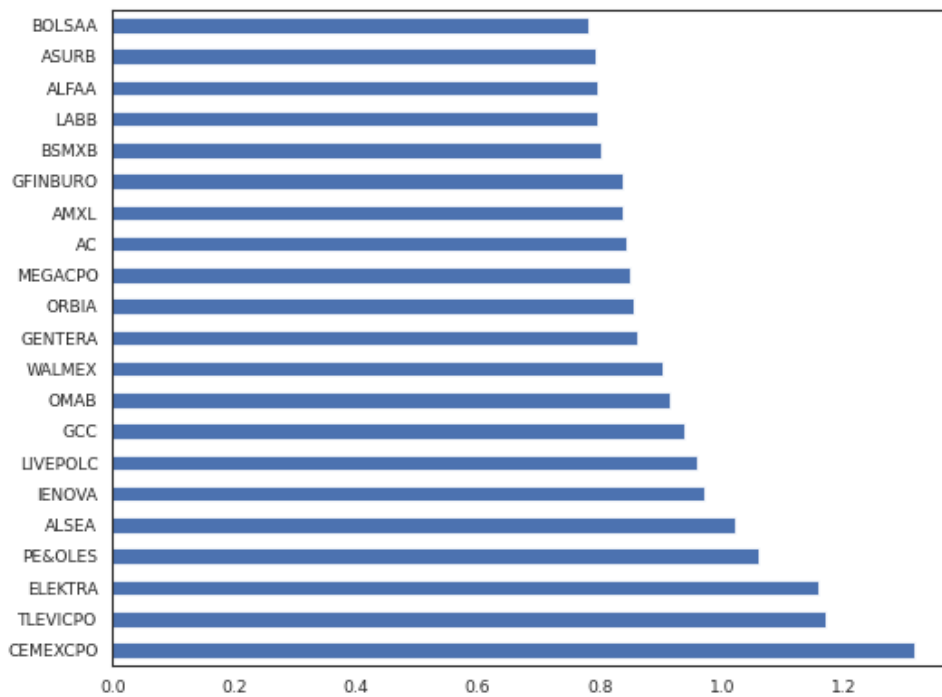


Figura 4.2: Series accionarias con mayor información

Fuente: Elaboración propia.

La muestra final con la que se realizaron los experimentos se muestra en el cuadro 4.1.

## 4.2. Arquitectura del Modelo

La arquitectura general del modelo (Fig. 4.3) involucra cuatro etapas principales: (1) preprocesamiento de datos utilizando el filtro Hodrick-Prescott para descomponer las series en tendencia, ciclo y residuo; (2) aplicación de una capa convolucional para extraer las interacciones de corto plazo entre las series; (3) dos capas apiladas de redes neuronales recurrentes con memoria a corto y largo plazo (LSTM) para generar el pronóstico de los valores de tendencia, ciclo y residuo de la serie del IPC; (4) recomposición de la serie de IPC sumando los tres pronósticos.

### 4.2.1. Descomposición en tendencia, ciclo y residuo

La descomposición de series macroeconómicas puede realizarse mediante cualquiera de los enfoques generales del procesamiento de señales, como el análisis de Fourier o la

Cuadro 4.1: Muestra del IPC

#	Razón Social	Clave de Cotización
1	Arca Continental, S.A.B. de C.V.	AC
2	Alfa, S. A. B. de C. V.	ALFAA
3	Alsea, S.A.B. de C.V.	ALSEA
4	América Móvil, S.A.B. de C.V.	AMXL
5	Grupo Aeroportuario del Sureste, S. A. B. de C. V.	ASURB
6	Bolsa Mexicana de Valores, S.A.B. de C.V.	BOLSAA
7	CEMEX, S.A.B. de C.V.	CEMEXCPO
8	Grupo Elektra, S.A.B. de C.V.	ELEKTRA
9	Genera, S.A.B. de C.V.	GENTERA
10	Grupo Cementos de Chihuahua S.A.B. de C.V.	GCC
11	Grupo Financiero Inbursa, S.A.B. de C.V.	GFINBURO
12	Infraestructura Energética Nova, S.A.B. de C.V.	IENOVA
13	Genomma Lab Internacional, S.A.B. de C.V.	LABB
14	El Puerto de Liverpool, S.A.B. de C.V.	LIVEPOLC-1
15	Megacable Holdings, S. A. B. de C. V.	MEGACPO
16	Grupo Aeroportuario del Centro Norte, S.A.B. de C.V.	OMAB
17	Orbia Advance Corporation, S.A.B. de C.V.	ORBIA
18	Industrias Peñoles, S.A.B. de C.V.	PEÑOLES
19	Grupo Televisa, S.A.B.	TLEVISACPO
20	Wal-Mart de México, S.A.B. de C.V.	WALMEX

descomposición en *wavelets*. Sin embargo, las series macroeconómicas exhiben patrones que pueden explicarse desde la teoría económica. Este es el principio que fundamenta el filtro Hodrick-Prescott (Hodrick y Prescott, 1997), el cual sugiere que la extracción del componente cíclico se encuentra relacionado con la teoría de los ciclos económicos.

La idea principal parte del supuesto de que toda serie de tiempo macroeconómica se encuentra formada por la suma de dos factores: el componente tendencial  $g_t$  y el componente cíclico  $c_t$ , es decir:

$$y_t = g_t + c_t \quad (4.1)$$

donde  $t = 1, 2, 3, \dots, T$ .

Desde este punto de vista, la extracción del componente tendencial se reduce a la solución del siguiente problema de minimización:

$$\min_{g_t} = \left\{ \sum_{t=1}^T C_t^2 + \lambda \sum_{t=2}^{T-1} [(g_{t+1} - g_t) - (g_t - g_{t-1})]^2 \right\} \quad (4.2)$$

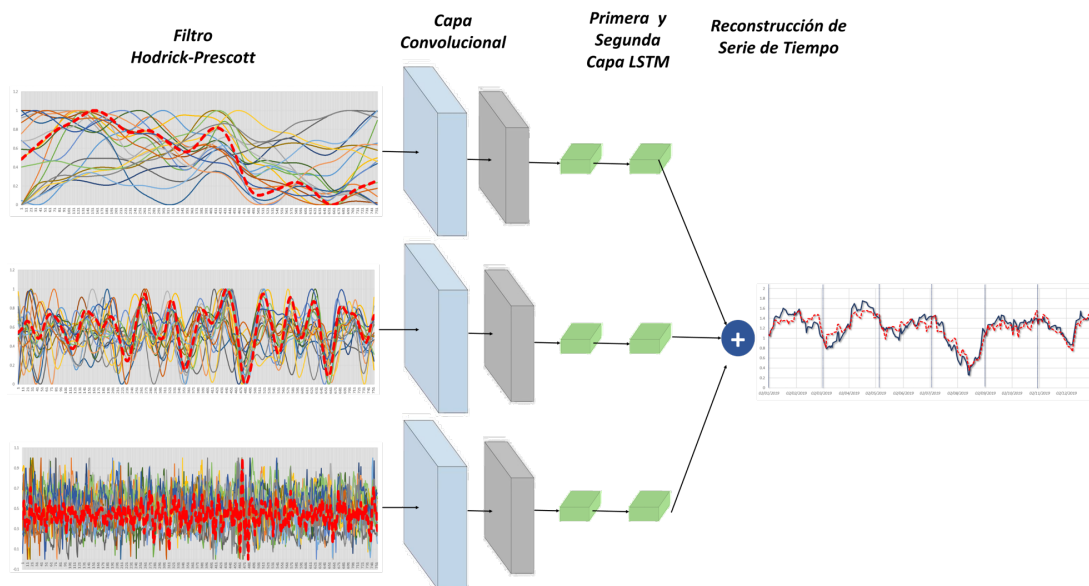


Figura 4.3: Arquitectura propuesta

Fuente: Elaboración propia.

El primer componente de la ecuación 4.2 corresponde a las diferencias entre la serie original y la tendencia, mientras que el segundo componente es la medida de suavizamiento de la serie.

Por lo tanto, el parámetro  $\lambda$  (conocido como *parámetro de suavización*) sirve como factor de penalización de la variabilidad en la tendencia. Puede observarse que si  $\lambda \rightarrow 0$  el componente tendencial será igual a la serie original  $y_t$ , mientras que si  $\lambda \rightarrow \infty$  la tendencia se convertirá en una serie completamente lineal.

En la mayoría de los trabajos que analizan datos macroeconómicos como PIB o empleo, el parámetro  $\lambda$  toma un valor de 1,600 para datos que tienen una frecuencia trimestral y de 14,400 para datos mensuales (Morten O. y Harald, 2002).

Sin embargo, Kaiser y Maravall (2001) sugieren seleccionar el parámetro en función del periodo de actividad cíclica que se desea analizar, es decir, en función del corte de las frecuencias bajas.

$$\lambda = \frac{1}{4 \left[1 - \cos\left(\frac{2\pi}{T}\right)\right]^2} \quad (4.3)$$

Donde  $T$  es el tamaño del ciclo de referencia sobre el que se desea hacer el estudio. Para el presente estudio y dado que los datos muestreados tienen una frecuencia diaria se

utilizó el parámetro  $T = 120$ , dando como resultado un valor de  $\lambda = 129,600$ .

Después de aplicar el filtro, a la componente cíclica se le hace pasar nuevamente por el filtro con los mismos parámetros, obteniendo tres series de tiempo a las que definimos como tendencia (Figura 4.4b), ciclo suavizado (Figura 4.4c) y residuo (Figura 4.4d).

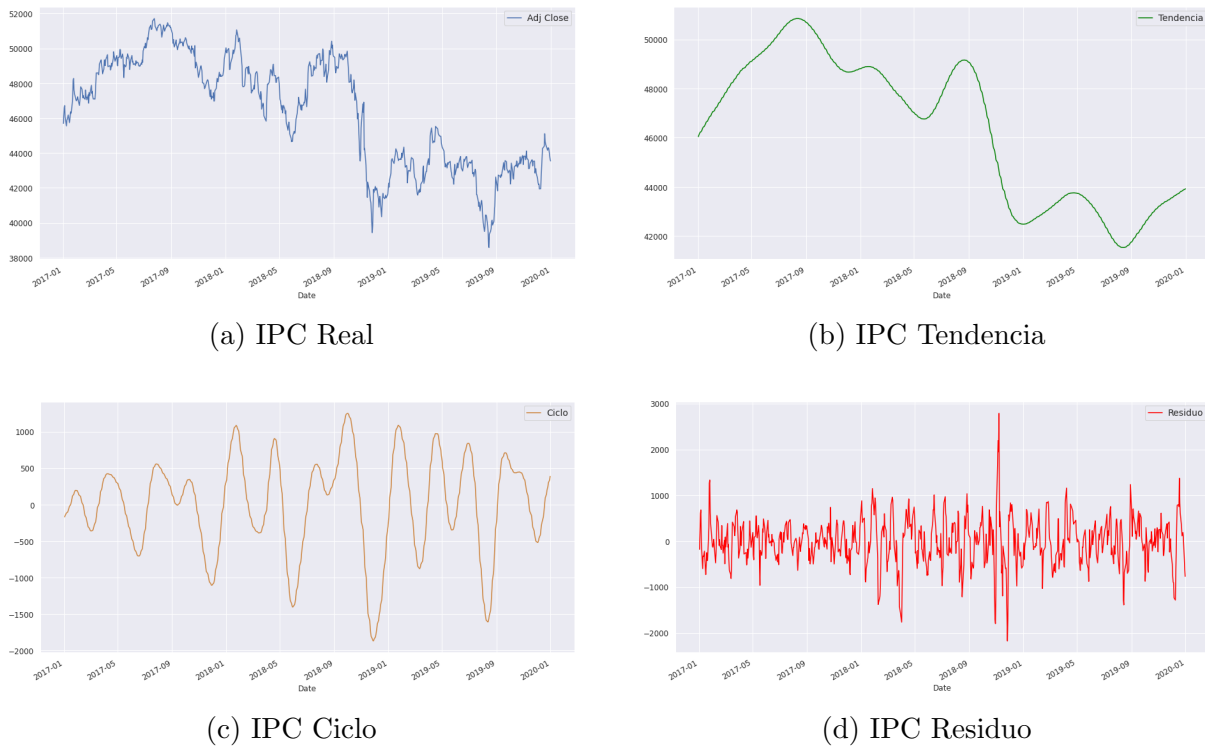


Figura 4.4: Descomposición de la serie en tendencia, ciclo y residuo

Fuente: Elaboración propia.

El mismo procedimiento de filtrado se aplicó a todas las series accionarias utilizando los mismos parámetros. Las series obtenidas se normalizan y se agruparon en matrices de tamaño  $753 \times 20$  que representan las entradas al modelo y tres matrices de tamaño  $753 \times 1$  que son los datos objetivo a pronosticar (Figuras 4.5, 4.6, 4.7).

### 4.2.2. Capa Convolutiva

Las redes neuronales convolucionales se desarrollaron con la idea de conectividad local: cada nodo está conectado a una región local en la entrada. La conectividad local se logra reemplazando las sumas ponderadas de la red neuronal tradicional con operaciones de convolución, como se observó en la sección 3.3.3. En cada capa de la red neuronal, la entrada es convolucionada con una matriz de pesos (también llamada filtro) para crear

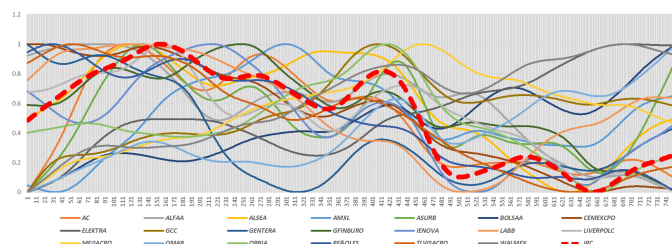


Figura 4.5: Matriz de tendencias

Fuente: Elaboración propia.

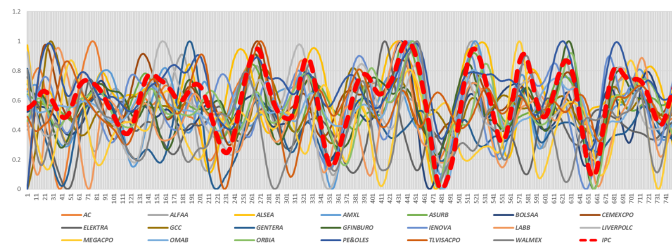


Figura 4.6: Matriz de ciclos

Fuente: Elaboración propia.

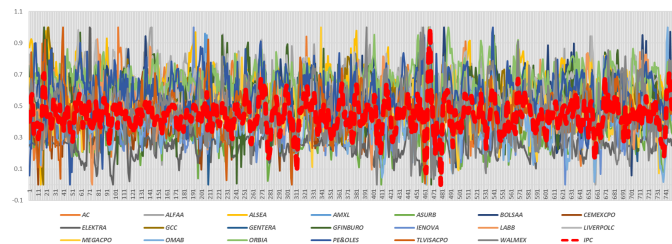


Figura 4.7: Matriz de residuos

Fuente: Elaboración propia.

un mapa de características, que representa las interacciones entre las diferentes series. La matriz de pesos (el filtro) se desliza sobre la entrada y calcula la convolución entre la matriz de entrada y los pesos del filtro.

A diferencia de las redes neuronales tradicionales, todas las neuronas que generan los valores en el mapa de características de salida comparten los mismos pesos. Esto significa que todos los nodos en la salida detectan el mismo patrón. La conectividad local y los pesos compartidos de las CNN reducen el número total de parámetros que se pueden aprender, lo que resulta en un aprendizaje más eficiente. La intuición detrás de una red neuronal convolucional es, por lo tanto, aprender en cada capa los pesos adecuados de los filtros que serán capaces de extraer las características de la entrada que relacionan los

puntos vecinos dentro del tamaño del filtro.

El parámetro  $k$  que representa el tamaño de filtro controla el campo receptivo de cada nodo. Sin *zero-padding*, la salida de convolución en cada capa tiene un ancho  $N_l = N_{l-1} - k + 1$ . Dado que todos los elementos en el mapa de características comparten los mismos pesos, esto permite que las características se detecten de manera invariante en el tiempo. La salida de la red después de  $L$  capas convolucionales será la matriz de características  $f_L$ , cuyo tamaño depende del tamaño del filtro y del número de filtros utilizados en la capa final.

En el presente trabajo se utilizó una sola capa convolucional con 64 filtros de tamaño 1x4 sin *zero-padding* y con función de reducción Max-Pooling de tamaño 2. La función de activación aplicada a la salida es la función *ReLU*.

### 4.2.3. Capas Recurrentes LSTM

La información obtenida de la capa de red convolucional es alimentada como entrada a una de red neuronal recurrente. La capa recurrente está constituida por unidades de memoria de corto y largo plazo (LSTM).

El funcionamiento general de la capa recurrente queda descrita por:

$$\begin{aligned}
 f_{(t)} &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned} \tag{4.4}$$

Estas ecuaciones representan las distintas compuertas que implementa una unidad LSTM, una compuerta de olvido  $f_{(t)}$ , una compuerta de entrada  $i_{(t)}$ , una compuerta de actualización  $C_{(t)}$  y una de salida  $o_{(t)}$ .

El valor de salida y el estado actual de la unidad LSTM permiten calcular un valor de pronóstico  $h_{(t)}$ .

Para la capa recurrente se utilizaron dos capas LSTM, la primera con 100 unidades y la segunda con 50 unidades, todas con las funciones de activación estándar.



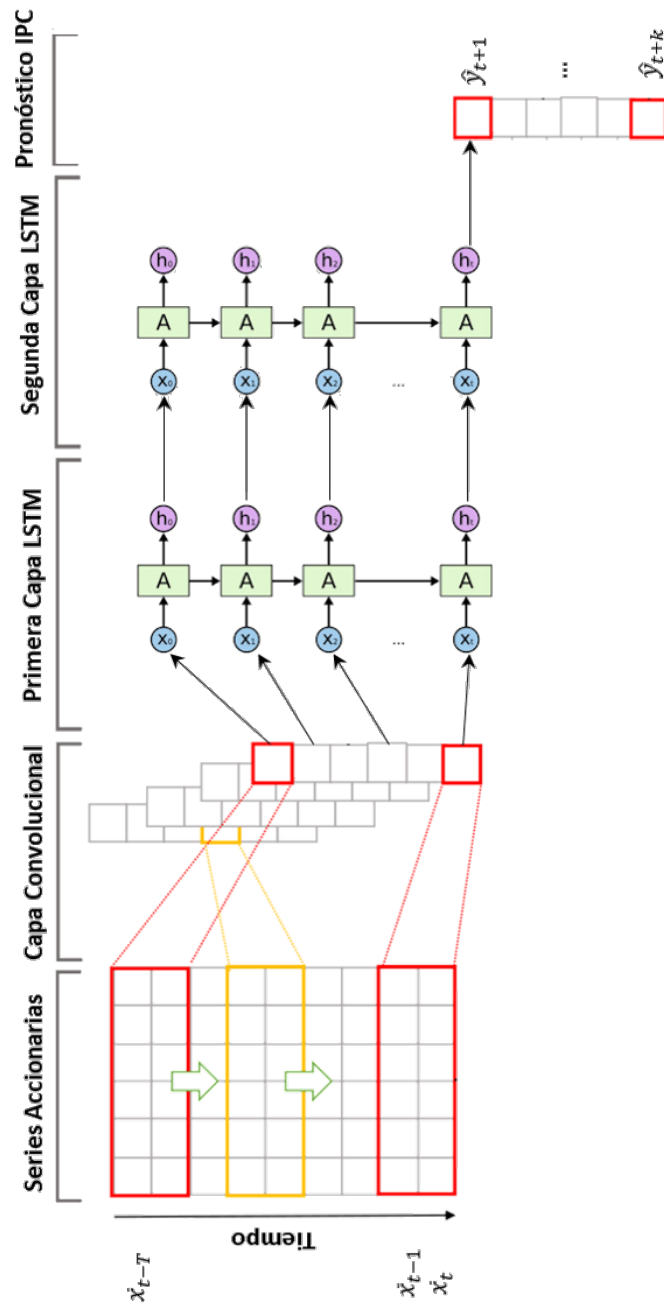


Figura 4.8: Arquitectura de aprendizaje profundo

Fuente: Elaboración propia.

# Capítulo 5

## Resultados

### 5.1. Experimentos

El procedimiento de predicción consta de tres partes. La primera es la parte de entrenamiento de la red, que se utiliza para que el modelo analice los datos de entrada y actualice los parámetros de sus pesos sinápticos. La segunda es la parte de prueba. Se utiliza para ajustar los hiperparámetros y obtener una configuración de modelo óptima. La última etapa es la validación del modelo, donde se realizan los pronósticos finales a evaluar.

Los datos se dividen de acuerdo al estándar en la mayoría de los estudios relacionados con el aprendizaje de máquinas, donde se recomienda utilizar aproximadamente el 80 % de las observaciones para la etapa de entrenamiento, 10 % de las observaciones para la etapa de pruebas y 10 % para la etapa de validación. En el presente trabajo para la etapa de entrenamiento se utilizan los datos de los dos años para capacitar al modelo (84 %), los siguientes dos meses (7 %) se emplean en la etapa de prueba. Finalmente, en la etapa de validación, se predicen los valores del IPC para los siguientes dos meses (7 %). Esta ventana se va recorriendo dos meses, lo que permite pronosticar los valores para todo el año 2019.

El procedimiento de predicción se ilustra en la Figura 5.1. Los detalles sobre los seis intervalos empleados pueden observarse en la Tabla 5.1

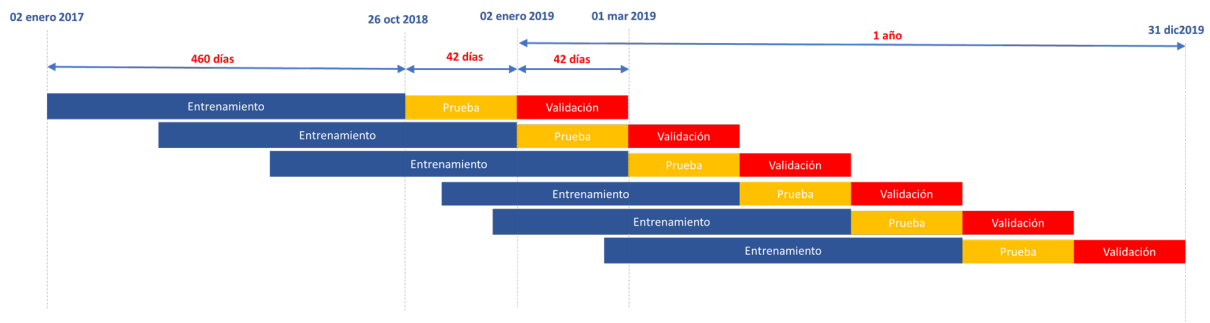


Figura 5.1: Segmentación de datos

Fuente: Elaboración propia.

Cuadro 5.1: Periodos

	<i>Entrenamiento</i>	<i>Prueba</i>	<i>Validación</i>
<b>Periodo 1</b>	02/01/2017 al 26/10/2018	29/10/2018 al 31/12/2018	02/01/2019 al 01/03/2019
<b>Periodo 2</b>	02/03/2017 al 31/12/2018	02/01/2019 al 01/03/2019	04/03/2019 al 06/05/2019
<b>Periodo 3</b>	05/05/2017 al 01/03/2019	04/03/2019 al 06/05/2019	07/05/2019 al 03/07/2019
<b>Periodo 4</b>	04/07/2017 al 06/05/2019	07/05/2019 al 03/07/2019	04/07/2019 al 30/08/2019
<b>Periodo 5</b>	31/08/2017 al 03/07/2019	04/07/2019 al 30/08/2019	02/09/2019 al 30/10/2019
<b>Periodo 6</b>	30/10/2017 al 30/08/2019	02/09/2019 al 30/10/2019	31/10/2019 al 31/12/2019

Para el periodo de validación, el IPC presentó el siguiente comportamiento real (Fig. 5.2). En el segundo bimestre mostró su mejor nivel, alcanzando el máximo de ese año. El tercer bimestre es de una enorme estabilidad pero que precede a una fuerte caída durante el cuarto bimestre, aunque presenta cierta recuperación antes del cierre del mismo bimestre. Los últimos dos bimestres mantienen esta recuperación y observamos el índice en los mismo niveles que al inicio de año.



Figura 5.2: Precio de cierre ajustado del IPC en 2019

Fuente: Elaboración propia.

Se realizaron los pronósticos para los seis periodos de prueba utilizando el modelo

propuesto. En la gráfica 5.3 se presenta la serie real de cotizaciones en comparación con el pronóstico extraído del modelo propuesto (CNN-LSTM), basado en una arquitectura de aprendizaje profundo. En la Figura 5.4 se observa el grado de acierto alcanzado por el pronóstico de aprendizaje profundo contra la cotización real.

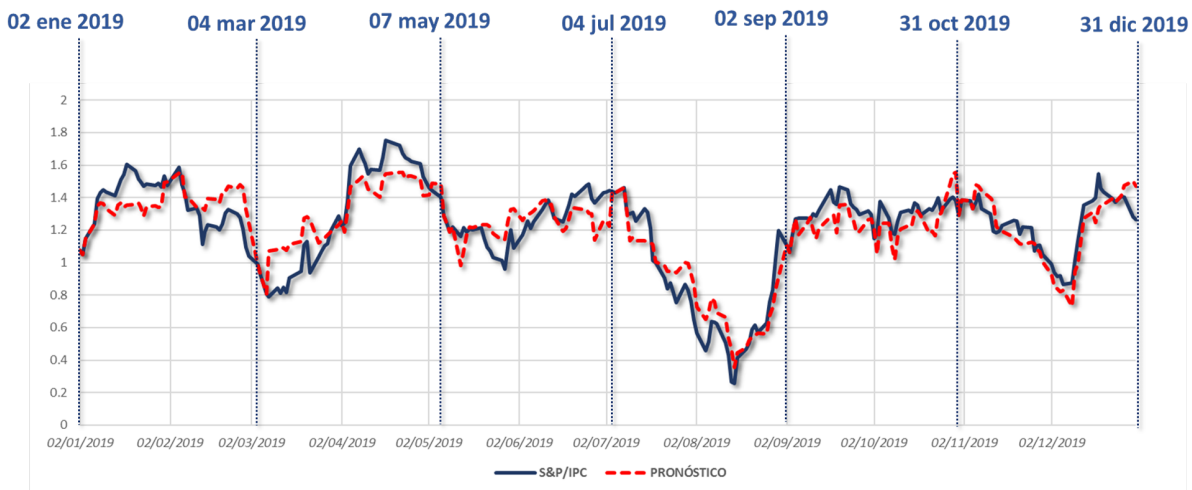


Figura 5.3: Real y pronóstico CNN-LSTM para los periodos de 2019

Fuente: Elaboración propia.

Adicionalmente, se hicieron pronósticos para los mismos periodos con un modelo univariado autorregresivo (AR) tomando como dato de entrada la serie del IPC y un modelo multivariado de vectores autorregresivos (VAR) utilizando las 20 series accionarias representativas. Los resultados comparativos de los tres modelos se muestran en la Figura 5.4. Puede observarse que el grado de acierto del modelo de aprendizaje profundo es mayor y más estable que el de los modelos econométricos tradicionales (AR y VAR) al ser comparados con la cotización real.

## 5.2. Evaluación de la precisión predictiva

En esta sección se presentan las métricas para evaluar el rendimiento del modelo. Para este trabajo se eligieron cuatro métricas tradicionales utilizadas para evaluar modelos generales de aprendizaje automático (MAPE, MSE y RMSE) así como la capacidad predictiva en series de tiempo financieras (Thiel U).

Las definiciones de los indicadores son las siguientes:

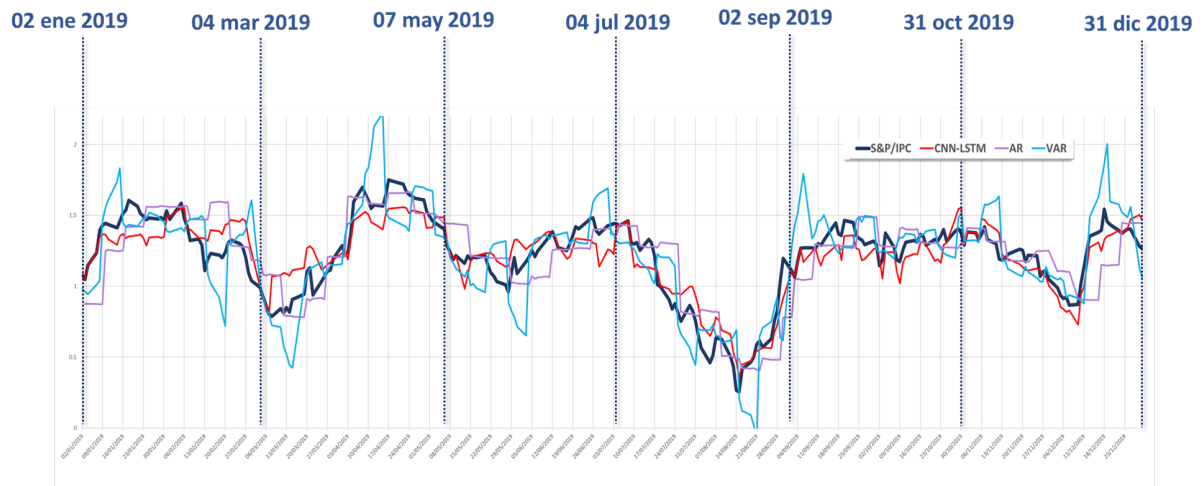


Figura 5.4: Pronósticos para los periodos de 2019

Fuente: Elaboración propia.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (5.1)$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (5.2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2} \quad (5.3)$$

$$Thiel U = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t)^2 + \frac{1}{N} \sum_{t=1}^N (\hat{y}_t)^2}} \quad (5.4)$$

En todas estas ecuaciones,  $y_t$  es el valor real mientras que  $\hat{y}_t$  es el valor predicho.  $N$  representa la cantidad de días en el período de predicción.

Los valores MAPE (Error Medio de Porcentaje Absoluto), MSE (Error Cuadrático Medio), y RSE (Raíz del Error Cuadrático Medio) miden el tamaño del error, mientras que Thiel U es una medida relativa de la diferencia entre dos variables.

En todos los casos, si los resultados son pequeños indican que el valor predicho está cerca del valor real. Los resultados comparativos para cada una de las métricas se encuen-

tran en los cuadros 5.2, 5.3, 5.4 y 5.5, resaltando en negrillas los mejores resultados para cada periodo.

Tanto en los casos MAPE, MSE y RMSE, el modelo CNN-LSTM presenta un mejor resultado que el de las metodologías AR y VAR en todos los periodos excepto el segundo, lo cual se explica debido a la menor volatilidad presentada en ese periodo. Es en el cuarto periodo de todas las métricas cuando se presenta el error más alto, el cual, no obstante, es menor que los presentados en las metodologías AR y VAR

Cuadro 5.2: Precisión Predictiva: MAPE

<i>Modelo</i>	<i>Periodo 1</i>	<i>Periodo 2</i>	<i>Periodo 3</i>	<i>Periodo 4</i>	<i>Periodo 5</i>	<i>Periodo 6</i>	<i>Promedio</i>
CNN-LSTM	<b>0.099436</b>	0.116293	<b>0.085099</b>	<b>0.178277</b>	<b>0.073102</b>	<b>0.071424</b>	<b>0.103939</b>
AR	0.144142	<b>0.093111</b>	0.101147	0.269293	0.078684	0.118272	0.134108
VAR	0.129450	0.130458	0.117153	0.307167	0.109236	0.111303	0.150794

Cuadro 5.3: Precisión Predictiva: MSE

<i>Modelo</i>	<i>Periodo 1</i>	<i>Periodo 2</i>	<i>Periodo 3</i>	<i>Periodo 4</i>	<i>Periodo 5</i>	<i>Periodo 6</i>	<i>Promedio</i>
CNN-LSTM	<b>0.021998</b>	0.023650	<b>0.015166</b>	<b>0.018972</b>	<b>0.011726</b>	<b>0.010622</b>	<b>0.017022</b>
AR	0.052412	<b>0.018797</b>	0.022712	0.057157	0.015238	0.034393	0.033451
VAR	0.047240	0.050135	0.034588	0.065425	0.030297	0.034470	0.043693

Cuadro 5.4: Precisión Predictiva: RMSE

<i>Modelo</i>	<i>Periodo 1</i>	<i>Periodo 2</i>	<i>Periodo 3</i>	<i>Periodo 4</i>	<i>Periodo 5</i>	<i>Periodo 6</i>	<i>Promedio</i>
CNN-LSTM	<b>0.148317</b>	0.153787	<b>0.123150</b>	<b>0.137738</b>	<b>0.108289</b>	<b>0.103063</b>	<b>0.129057</b>
AR	0.228937	<b>0.137101</b>	0.150704	0.239075	0.123442	0.185453	0.177452
VAR	0.217347	0.223909	0.185980	0.255783	0.174062	0.185662	0.207124

Cuadro 5.5: Precisión Predictiva: Thiel U

<i>Modelo</i>	<i>Periodo 1</i>	<i>Periodo 2</i>	<i>Periodo 3</i>	<i>Periodo 4</i>	<i>Periodo 5</i>	<i>Periodo 6</i>	<i>Promedio</i>
CNN-LSTM	<b>0.053535</b>	0.056974	<b>0.048785</b>	<b>0.082452</b>	<b>0.042020</b>	<b>0.042198</b>	<b>0.054327</b>
AR	0.083308	<b>0.051396</b>	0.059926	0.129725	0.047207	0.075054	0.074436
VAR	0.079526	0.083849	0.073746	0.144164	0.064881	0.072382	0.086425

Con base en estos resultados, se concluye que el rango de error del método CNN-LSTM es el más pequeño: en promedio, de apenas 10% en el caso del MAPE, de 2% en el caso del MSE y del 12% en el RMSE, lo que demuestra la ventaja que el pronóstico mediante el modelo que utiliza redes neuronales de aprendizaje profundo mantiene por encima de los métodos tradicionales

### 5.3. Evaluación de los rendimientos.

Es posible crear una estrategia de compra-venta de activos financieros basada en los resultados pronosticados de cada modelo. La idea es que, bajo la misma estrategia comercial, podemos determinar el modelo que puede obtener las mayores ganancias para un inversor.

La estrategia recomienda que los inversores compren cuando el valor previsto para el próximo período sea mayor que el valor real actual. Por el contrario, recomienda que los inversores vendan cuando el valor predicho sea menor que el valor real actual.

Matemáticamente, las señales de compra y venta pueden describirse como:

$$\text{Comprar} = \hat{y}_{t+1} > y_t \quad (5.5)$$

$$\text{Vender} = \hat{y}_{t+1} < y_t \quad (5.6)$$

Donde  $y_t$  representa el valor actual del IPC y  $\hat{y}_{t+1}$  representa el valor pronosticado para el día siguiente.

Finalmente, la rentabilidad de cada modelo puede calcularse como:

$$R = \sum_{t=1}^B \frac{y_{t+1} - y_t + (y_t \cdot C_B + y_{t+1} \cdot C_S)}{y_t} + \sum_{t=1}^C \frac{y_t - y_{t+1} + (y_{t+1} \cdot C_B + y_t \cdot C_S)}{y_t} \quad (5.7)$$

Donde  $B$  y  $C$  denotan el número total de días para comprar y vender, respectivamente. Adicionalmente,  $C_B$  y  $C_S$  son los costos de transacción para comprar y vender.

Dado que los costos de transacción son iguales, independientemente del modelo uti-

lizado, para los fines comparativos de este estudio podemos despreciarlos y por lo tanto consideramos  $C_B = C_S = 0$ . La ecuación 5.7 se simplifica bastante, resultando:

$$R = \sum_{t=1}^B \frac{y_{t+1} - y_t}{y_t} + \sum_{t=1}^C \frac{y_t - y_{t+1}}{y_t} \quad (5.8)$$

Utilizando la Ecuación 5.8 se obtienen los siguientes resultados para los seis periodos con cada uno de los modelo empleados.

Cuadro 5.6: Rendimientos

<i>Modelo</i>	<i>Periodo 1</i>	<i>Periodo 2</i>	<i>Periodo 3</i>	<i>Periodo 4</i>	<i>Periodo 5</i>	<i>Periodo 6</i>	<i>Promedio</i>
CNN-LSTM	<b>0.442562</b>	<b>1.141294</b>	<b>0.220207</b>	<b>0.613222</b>	<b>0.556152</b>	<b>0.551770</b>	<b>0.587535</b>
AR	-0.432215	0.459338	-0.195216	-0.234979	0.343758	-0.518750	-0.096344
VAR	-0.691981	0.356674	-0.053396	-1.221284	0.497316	0.468560	-0.107352

Con base en el Cuadro 5.6, se concluye que, al utilizar la metodología CNN-LSTM es posible obtener un rendimiento por inversión del 50% anual, mientras que las otras metodologías producen, por el contrario, pérdidas: del 20% con la AR y del 10% con la VAR.



# Capítulo 6

## Conclusiones y Trabajo Futuro

En esta sección se describen las conclusiones obtenidas del presente estudio al generar pronósticos del IPC como una serie de tiempo multivariada empleando la metodología de las redes neuronales profundas. También se presentan algunos conceptos que se podrían considerar para continuar este trabajo a futuro.

La tesis ha probado su hipótesis central al mostrar que los resultados, técnicos y financieros, obtenidos mediante un algoritmo de aprendizaje profundo superan los modelos econométricos tradicionales y facilitan el manejo de grandes volúmenes de información.

Esto queda plasmado en la Figura 5.4 de forma gráfica y en las tablas 5.2 a 5.4, que hacen comparativos entre los resultados obtenidos por la arquitectura propuesta y los valores obtenidos por los modelos base.

Puede observarse que el modelo propuesto presenta una mayor estabilidad y menor error en sus pronósticos, de acuerdo a diferentes métricas estándar. Esto conduce naturalmente a una mayor rentabilidad en sus decisiones de compra y venta. El enfoque empleado permite también conocer independientemente los pronósticos de tendencia y ciclo para el IPC (Fig. 6.1).

Otra de las principales ventajas de las redes neuronales convolucionales es que sus capas y los pesos de los filtros, arrojan resultados que en muchas ocasiones son interpretables en el dominio del problema, por lo que en un futuro se planea desarrollar el tema del análisis individual de los pronósticos de las matrices de series accionarias, ya que facilitaría el manejo de un portafolio que replique el comportamiento del IPC.

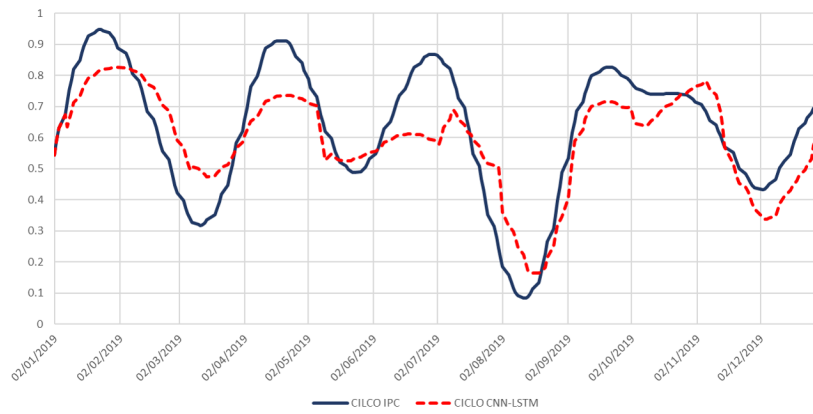


Figura 6.1: Ciclo real y pronóstico CNN-LSTM para 2019

Fuente: Elaboración propia.

Por otra parte, en fechas recientes, han aparecido artículos utilizando datos secuenciales que sugieren que las redes neuronales recurrentes pueden ser sustituidas por mecanismos de atención (Vaswani y cols., 2017), mejorando los resultados de pronóstico, particularmente para datos secuenciales que no son cíclicos, y otorgando una mayor interpretabilidad a los resultados. En nuestro modelo esto implicaría saber cuales periodos de tiempo previos han sido los más significativos en la predicción actual.

# Referencias

- Ablad, M., Frikh, B., y Ouhbi, B. (2020). Uncertainty Quantification in Deep Learning Context: Application to Insurance. *6th IEEE Congress on Information Science and Technology (CiSt) Information Science and Technology (CiSt)*.
- Arango, A., y Velásquez, J. D. (2014). Forecasting the colombian exchange market index (IGBC) using neural networks. *IEEE Latin America Transactions*, 12(4), 718–724.
- Arel, I., Rose, D. C., y Karnowski, T. P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research. (November), 13–18.
- Arrow, K. J. (1964). The role of securities in the optimal allocation of risk bearing. *The Review of Economic Studies*, 91–96.
- Atsalakis, G. S., y Valavanis, K. P. (2009). Surveying stock market forecasting techniques - Part II: Soft computing methods. *Expert Systems with Applications*, 36(3 PART 2), 5932–5941.
- Awad, M., y Khanna, R. (2015). *Efficient Learning Machines*. Berkeley, CA: Apress.
- Babu, A. M., y Pratap, A. (2020). Credit Card Fraud Detection Using Deep Learning. *2020 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2020*, 32–36.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques De L'É.N.S.*, 17, 21–86.
- Bengio, Y., Courville, A., y Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.
- Bengio, Y., Lamblin, P., Popovici, D., y Larochelle, H. (2007). Greedy layer-wise training

- of deep networks. *Advances in Neural Information Processing Systems*(1), 153–160.
- Bengio, Y., Simard, P., y Frasconi, P. (2014). Learning Long-term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Network*, 5(2), 157.
- Bernstein, P. L. (1996). *Against the Gods. The Remarkable Story of Risk*. New York: John Wiley & Sons, Inc.
- BMV, y S&P Dow Jones Índices. (2017). *Bolsa Mexicana de Valores (BMV) y S&P Dow Jones Indices (S&P DJI) anuncian cambios en la metodología de los índices de BMV* (Inf. Téc.).
- Box, G., Jenkins, G., y Reinsel, G. (2008). *Time series analysis*. New Jersey: Wiley.
- Britz, D. (2015). *Implementing a neural network from scratch in python – an introduction*. <http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/>.
- Cabrera, A., y Ortiz, F. (2012). Pronóstico del rendimiento del IPC (Índice de Precios y Cotizaciones) mediante el uso de redes neuronales diferenciales. *Contaduría y administración*, 57(2), 63–81.
- Cao, J., y Wang, J. (2019). Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12), 1–13.
- Cassese, G. (2020). Complete and Competitive Financial Markets in a Complex World. *SSRN Electronic Journal*, 1–21.
- Chen, X. W., y Lin, X. (2014). Big data deep learning: Challenges and perspectives. *IEEE Access*, 2, 514–525.
- Cootner, P. H. (1968). The Random Character of Stock Market Prices. *The Econometric Society*, 36(1), 191–192.
- Cowpertwait, P., y Metcalfe, A. (2011). *Introductory time series with R*. Springer.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2, 303–314.
- Dávila, Á., Sanchez-Choez, N., y Román-Vásquez, J. L. (2017). Pronóstico del Índice Bursátil Ecuatoriano (ECUINDEX) mediante redes neuronales autorregresivas. *Empresa*, 6, 16–32.

- Donoho, D. L. (2000). High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. *American Math. Society Lecture-Math Challenges of the 21st Century*(January 2000), 1–33.
- Fama, E. F. (1965). The Behaviour Of Stock-Market Prices. *The Journal of Business*, 34–105.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 421.
- Ferrari, R. (2005). La complejidad: un nuevo enfoque de la economía financiera. *Contaduría y Administración*(216), 73–100.
- Gantz, J., y Reinsel, D. (2010). The Digital Universe Decade – Are You Ready? , *2009*(May 2010), 1–16.
- Georgieva, S. (2016). Efficient Markets Hypotesis And Market Anomalies. *Entrepreneurship*, 34–47.
- Hartley, T. (2014). *When parallelism gets tricky: Accelerating floyd-steinberg on the mali gpu.* <https://community.arm.com/developer/tools-software/graphics/b/blog/posts/when-parallelism-gets-tricky-accelerating-floyd-steinberg-on-the-mali-gpu/>.
- Hawaldar, I. T., Rohit, B., y Pinto, P. (2017). Testing of weak form of efficient market hypothesis: Evidence from the Bahrain bourse. *Investment Management and Financial Innovations*, 14(2), 376–385.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. McMaster University, Ontario Canada: Pearson.
- Heaton, J. B., Polson, N. G., y Witte, J. H. (2016). Deep Learning in Finance. (February), 1–20.
- Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence*, 40(1-3), 185–234.
- Hochreiter, S., y Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- Hodrick, R. J., y Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical

- Investigation. *Journal of Money, Credit and Banking*, 29(1), 1.
- Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359–366.
- Hu, Y. J., y Lin, S. J. (2019). Deep Reinforcement Learning for Optimizing Finance Portfolio Management. *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 14–20.
- Kharin, Y. (2013). *Robustness in Statistical Forecasting*. Switzerland: Springer.
- Kuhlmann, M. (2014). Explaining financial markets in terms of complex systems. *Philosophy of Science*, 81(5), 1117–1130.
- Lecun, Y., Bengio, Y., y Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- LeCun, Y., Haffner, P., Bottou, L., y Bengio, Y. (1998). Object Recognition with Gradient-Based Learning. *Proceedings of the IEEE*(0).
- León, L., Landassuri, V., Orozco, H., y Quintana, M. (2018). Predicción del IPC mexicano combinando modelos econométricos e inteligencia artificial. *Revista Mexicana de Economía y Finanzas*, 13(4), 27.
- Li, G., Xiao, M., y Guo, Y. (2019). Application of Deep Learning in Stock Market Valuation Index Forecasting. *10th International Conference on Software Engineering and Service Science*, 551–554.
- Li, X., Li, Y., Yang, H., Yang, L., y Liu, X.-Y. (2019). *DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News* (n.º NeurIPS).
- Lipton, Z. C., Berkowitz, J., y Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. , 1–38.
- López de Prado, M. (2018). *Advances in Financial Machine Learning*. New Jersey: John Wiley & Sons, Inc.
- López Herrera, F. (2004). Modelado de la volatilidad y pronóstico del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores. *Contaduría y Administración*(213), 43–72.
- McCulloch, W. S., y Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.

- Minsky, M., y Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Massachusetts: MIT Press.
- Modugno, M. (2011). *NowCasting Inflation Using High-Frequency Data*. Frankfurt.
- Moller, B., y Reute, U. (2007). *Uncertainty Forecasting in Engineering*. Dresden: Springer.
- Morten O., R., y Harald, U. (2002). On adjusting the hodrick-prescott filter for the frequency of observations. *The Review of Economics and Statistics*, 84(2), 371.
- Newell, A. (1982). The knowledge level. *Artificial Intelligence*, 18(1), 87–127.
- Niu, T., Wang, J., Lu, H., Yang, W., y Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications*, 148, 113237.
- Okina, K., y Shiratsuka, S. (1999). Financial Market Globalization: Present and Future. *Monetary and Economic Studies*.
- Olah, C. (2015). *Understanding lstm networks*. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Pryima, S., Vovk, R., y Vovk, V. (2019). Using Artificial Neural Networks to Forecast Stock Market Indices. *2019 11th International Scientific and Practical Conference on Electronics and Information Technologies, ELIT 2019 - Proceedings*, 108–112.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.
- Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Russel, S., y Norvig, P. (2010). *Artificial intelligence: a modern approach*. Pearson Education.
- Sagir, A. M., y Sathasivan, S. (2017). The use of artificial neural network and multiple linear regressions for stock market forecasting. *Matematika*, 33(1), 1.
- Shubhendu, S., y Vijay, J. (2013). Applicability of Artificial Intelligence in Different Fields of Life. *International Journal of Scientific Engineering and Research (IJSER)*, 1(1), 28–35.
- S&P Dow Jones Índices. (2014). Index Mathematics - Methodology. *Mc Graw Hill*

- Financial*(March), 62.
- Tetlock, P. E., y Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. New York: Crown Publishers.
- Turing, A. (1950). Computing Machinery And Intelligence. *Mind*, 49(5), 433–460.
- Vanstone, B., y Finnie, G. (2009). An empirical methodology for developing stockmarket trading systems using artificial neural networks. *Expert Systems with Applications*, 36(3 PART 2), 6668–6680.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Dec*, 5999–6009.
- Wang, J. Z., Wang, J. J., Zhang, Z. G., y Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, 38(11), 14346–14355.
- Wu, B., y Duan, T. (2019). Nonlinear Dynamics Characteristic of Risk Contagion in Financial Market Based on Agent Modeling and Complex Network. *Complexity*, 2019.





# Apéndice A

## Muestra del IPC

Cuadro A.1: Muestra del IPC

#	<i>Razón Social</i>	<i>Clave de Cotización</i>
1	Arca Continental, S.A.B. de C.V.	AC
2	Alfa, S. A. B. de C. V.	ALFAA
3	Alpek S.A.B. de C.V.	ALPEKA
4	Alsea, S.A.B. de C.V.	ALSEA
5	América Móvil, S.A.B. de C.V.	AMXL
6	Grupo Aeroportuario del Sureste, S. A. B. de C. V.	ASURB
7	Banco del Bajío, S.A., Institución de Banca Multiple	BBAJIOO
8	Grupo Bimbo, S.A.B. de C.V.	BIMBOA
9	Bolsa Mexicana de Valores, S.A.B. de C.V.	BOLSAA
10	Banco Santander Mexico, S.A.	BSMXB
11	CEMEX, S.A.B. de C.V.	CEMEXCPO
12	Becele, S.A.B. de C.V.	CUERVO
13	Grupo Elektra, S.A.B. de C.V.	ELEKTRA
14	Fomento Económico Mexicano, S.A.B. de C.V.	FEMSAUBD
15	Grupo Aeroportuario del Pacífico, S.A.B. de C.V.	GAPB
16	Grupo Carso, S.A.B. de C.V.	GCARSOA1
17	Grupo Cementos de Chihuahua S.A.B. de C.V.	GCC
18	Genera, S.A.B. de C.V.	GENTERA
19	Grupo Financiero Inbursa, S.A.B. de C.V.	GFINBURO
20	Grupo Financiero Banorte, S.A.B. de C.V.	GFNORTEO
21	Grupo Mexico, S.A.B. de C.V.	GMEXICOB
22	Gruma, S.A.B. de C.V.	GRUMAB
23	Infraestructura Energética Nova, S.A.B. de C.V.	IENOVA
24	Kimberly-Clark de Mexico, S. A. B. de C. V.	KIMBERA
25	Coca-Cola FEMSA, S.A.B. de C.V.	KOFUBL
26	Genomma Lab Internacional, S.A.B. de C.V.	LABB
27	El Puerto de Liverpool, S.A.B. de C.V.	LIVEPOLC-1
28	Megacable Holdings, S. A. B. de C. V.	MEGACPO
29	Grupo Aeroportuario del Centro Norte, S.A.B. de C.V.	OMAB
30	Orbia Advance Corporation, S.A.B. de C.V.	ORBIA
31	Industrias Peñoles, S.A.B. de C.V.	PE&OLES
32	Promotora y Operadora de Infraestructura, S. A. B. de C. V.	PINFRA
33	Regional S.A.B. de C.V.	RA
34	Grupo Televisa, S.A.B.	TLEVISACPO
35	Wal-Mart de México, S.A.B. de C.V.	WALMEX

# Apéndice B

## Matriz de Correlación

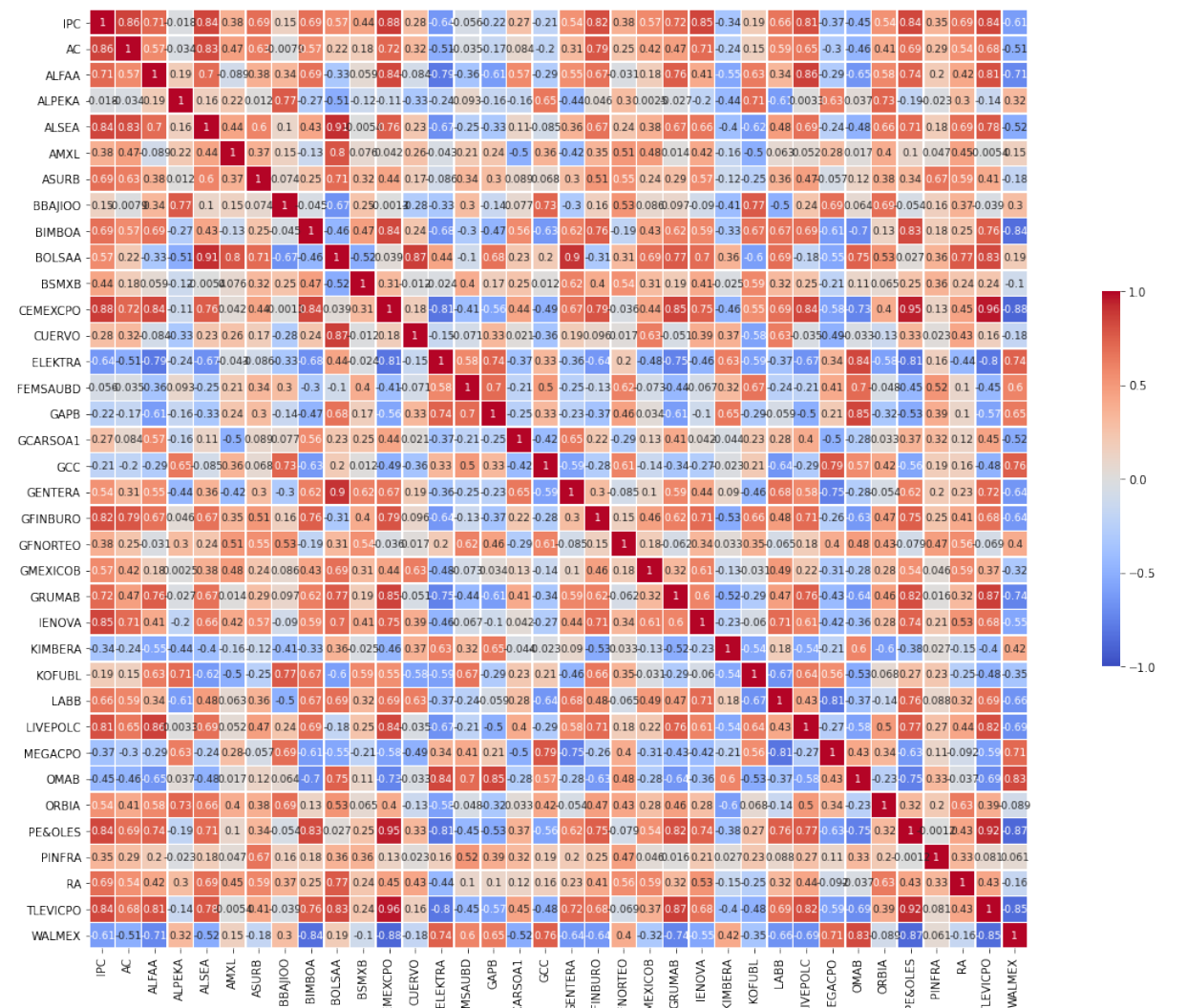


Figura B.1: Matriz de Correlación

Fuente: Elaboración propia.



