

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY



ESCUELA DE GOBIERNO Y TRANSFORMACIÓN PÚBLICA

**PRONÓSTICO DE PRECIOS DE PETRÓLEO PARA LA MEZCLA MEXICANA USANDO REDES
NEURONALES ARTIFICIALES**

TESIS

PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL GRADO ACADÉMICO DE:

MAESTRO EN ECONOMÍA APLICADA

POR:

FRANCISCO NAGIB RUIZ REYES

CIUDAD DE MÉXICO, MÉXICO.

DICIEMBRE 2022

Dedicatoria

A Dios. *Eben-Ezer*. Hasta aquí nos ayudó Jehová.

A mi esposa por su inconmensurable apoyo.

A mi madre por haber resistido y cruzado el desierto.

A mi hermano por su amor y lealtad.

A mis abuelos porque comencé donde ellos terminaron.

Agradecimientos

Al Dr. Roberto Barrera por su tiempo y su incalculable apoyo.

Al Dr. Luis Serra por su guía en el presente trabajo.

Al Dr. José Antonio Núñez por su disposición y su amabilidad.

A mi familia. Su orgullo es mi fuerza.

A Dios por permitirme entender una infinitesimal parte de su gran diseño.

Y dediqué mi corazón a conocer la sabiduría, y también a entender las locuras y los desvaríos; conocí que aún esto era aflicción de espíritu.

Porque en la mucha sabiduría hay mucha molestia; y quien añade ciencia, añade dolor.

Eclesiastés 1:17-18. RVR 1960

Vanidad de vanidades, dijo el Predicador; vanidad de vanidades, todo es vanidad.

Eclesiastés 1:2. RVR 1960

Aviso de Responsabilidad

Por la presente declaro que este trabajo de tesina es de mi propia autoría. Todas las fuentes y ayudas utilizadas se han indicado como tales. Todos los textos citados directamente o parafraseados han sido indicados por citas en el texto. Se dan detalles bibliográficos completos, en la lista de referencia también contiene fuentes de internet, por tanto, asumo la responsabilidad por los errores del presente trabajo en caso de que hubiese.

Contenido

1. Introducción y planteamiento del problema	6
1.1 Introducción.....	6
1.2 Planteamiento del problema	8
2. Pregunta de investigación e hipótesis	10
2.1 Pregunta de Investigación	10
2.2 Hipótesis	11
3. Revisión de literatura y marco teórico.....	12
3.1 Revisión de Literatura	12
3.2 Marco Teórico	14
4. Metodología - especificar la metodología.....	24
4.1 Selección de Datos	24
4.2 Variables significativas.....	26
4.3 Procedimiento	28
5. Análisis y resultados.....	29
5.1 Análisis.....	29
5.2 Resultados.....	37
5.2.1 Modelo Base	37
5.2.2 Variables significativas usando el algoritmo F Test para modelos de <i>Machine Learning</i>	38
5.2.3 Redes Neuronales.....	40
6. Conclusiones y recomendaciones.....	45
6.1 Conclusiones.....	45
6.2 Recomendaciones.....	47

1. Introducción y planteamiento del problema

1.1 Introducción

¿Qué sería de nosotros si supiésemos lo que pudiera pasar en el futuro? ¿Si supiéramos la fecha de nuestra muerte cambiarías algo de lo que estás haciendo ahora? No es difícil imaginarnos que todos nos hemos hecho la misma pregunta ¿Qué nos deparará el mañana?

Así mismo, la vida nos hace preguntarnos si tiene en sí misma una naturaleza determinística o, por el contrario, todo recae en un mero concepto de aleatoriedad (estocástico), pues no podemos corromper la cadena de causa-consecuencia, i.e. que cualquier evento es el resultado de su condición precedente. Y si la física fuera determinística y universal ¿En dónde se encuentra el libre albedrío si ni siquiera tenemos la elección de las condiciones iniciales? (Bishop & beim Graben, 2016)

Si se recuerda la película de “Volver al Futuro”, en un episodio, el antagonista roba un almanaque deportivo que usa posteriormente para volverse rico. Tal pareciera que el director de la película quisiera mostrar la relación de las personas con el dinero como una conexión tan fuerte que, hasta podríamos considerarla como parte de una necesidad del ser humano, a tal grado de cometer acciones que no se considerarían éticas. En ese orden de ideas es donde podemos mencionar a (Oleson, 2004) quien relaciona a las personas con respecto a las actitudes con el dinero, sus necesidades y la demográfica, dicho autor concluye que la importancia del dinero, cuando se está en los niveles más bajos de la jerarquía de necesidades de Maslow, es más fuerte y al mismo tiempo se va debilitando conforme se sube en esa pirámide.

Entonces, si consideramos al dinero como una motivación importante para una persona que va escalando en la pirámide de Maslow, e inclusive lo hacemos de manera agregada en una sociedad, mediante corporativos, familias o empresas, saber cómo hacer crecer el patrimonio se convierte en una tarea esencial. Y de acuerdo con (Boshara & Rademacher, 2021) hay 4 formas de capital como podemos construir riqueza: uno es invirtiendo en capital humano, otro en capital de salud, además de capital empresarial y financiero, así como capital digital, aunque no por eso sabremos si nuestro capital está bien invertido, puesto que para ello necesitaríamos conocer el futuro de nuestra métrica, ya sea productividad, rendimiento o dinero.

Independientemente de que se pueda o no conocer el futuro, al menos podemos aspirar a entender las probabilidades como lo quería Louis Bachelier, según lo expresa (Weatherall, 2013), aunque no sería la primera vez que un científico desarrollara algún modelo que intentara explicar los comportamientos de los mercados bursátiles y sobre todo para sacarles provecho. Posiblemente eso pensaba Samuelson cuando por primera vez escuchó de la tesis de Bachelier, desarrollada en torno al 1900, en la cual se veían conceptos que habían revolucionado la materia o eran relativamente nuevos, como el proceso de Wiener, las ecuaciones de Kolmogorov o la martingala de Doob, según lo relata (Weatherall, 2013).

La tesis de Bachelier posiblemente pueda considerarse como el origen de la “financiera matemática moderna” y de varias ramas importantes de cálculo estocástico, como la

teoría del movimiento Browniano y los procesos de Markov de acuerdo con (Villalón, Rodríguez Ruiz, & Seijas Macías, 2014).

El modelo de Bachelier planteado en su tesis de *Théorie de la Spéculation* (en español Teoría de la especulación) se sigue usando a la fecha e inclusive compite con el modelo de *Black Scholes*¹, así lo plantea (Galeeva & Ronn, 2020) al querer modelar los precios y las volatilidades del mercado petrolero; un objeto ampliamente estudiado.

El mercado petrolero es un área que ha atraído a la investigación porque los abundantes suministros de petróleo forman las bases de las economías industriales modernas, de acuerdo con (Miller & Sorrell, 2013) quienes establecen la disyuntiva entre los comentaristas que pronostican un pico en el futuro cercano y una posterior disminución de la producción mundial de petróleo, mientras que otros destacan el crecimiento de la producción de petróleo por el “petróleo compacto”² y el desarrollo de recursos no convencionales para su extracción.

Si cambiamos la perspectiva de los párrafos anteriores a un esquema de bien colectivo, en este caso en un país como México, y tomamos al petróleo como nuestra variable de estudio, el hecho de conocer su precio nos serviría para proteger al país y darle margen de maniobra en la toma de decisiones, pues de acuerdo con (Ramones Fernández, 2014) del CIEP³, existen tres contraprestaciones a favor del estado que son afectadas directamente por el precio del petróleo: las regalías, la tasa aplicada al valor contractual de los hidrocarburos y el derecho de extracción de hidrocarburos, además si tomamos en cuenta a (Vázquez Pérez, 2020) igualmente del CIEP, quien menciona que a principios de 2020, los precios de petróleo tuvieron su mayor caída trimestral de la historia y las coberturas petroleras compensaron alrededor del 36% del monto total de los ingresos petroleros entonces las coberturas fungen como amortiguadores. Es por ello por lo que en el presente trabajo tomamos como ejercicio de estudio al pronóstico del precio de la mezcla mexicana de petróleo.

Dentro de los pronósticos, el área común de estudio se encuentra dentro de los modelos estadísticos y si bien, las redes neuronales artificiales usan también una base matemática, existe un enfoque diferente entre ellas. (Hill, Marquez, Marcus, & Remus, 1994) señalan que en 1993 existía la duda acerca de si las redes neuronales podrían ser mejor herramienta de pronóstico que los métodos estadísticos convencionales, y estos concluyen que, bajo ciertas condiciones, es verdad. Consecuentemente, usaremos los algoritmos de redes neuronales para pronosticar los precios del crudo en México.

¹ Modelo de valuación de opciones.

² Es un tipo de petróleo que se encuentra en depósitos impermeables de esquisto y roca caliza. Para extraerse se usa un tipo de facturación hidráulica llamada *fracking* en inglés.

³ Centro de Investigación Económica y Presupuestaria.

1.2 Planteamiento del problema

México posee recursos petroleros que han dado periodos de auge en su economía y, al mismo tiempo, periodos de crisis económica. Hoy en día, el país tiene una dependencia considerable de estos recursos en sus finanzas públicas, aunque cada vez menor, esto se puede ver en la publicación del (Instituto Belisario Domínguez, 2022) con el reporte trimestral de ingresos presupuestarios en donde los ingresos petroleros de enero a junio representaron en millones de pesos un valor observado de 594,193.7 de un total de 3,305,119.5, lo que representa un 17.98%; remontamos al mismo periodo pero del 2010, de acuerdo con la (Secretaría de Hacienda y Crédito Público, 2010), los ingresos petroleros representaban en millones de pesos un valor observado de 452,742.4 de un total de 1,409,279.7, lo cual representa un 32.13%. Hay que recordar que, en los últimos 60 años, los mejores para México, tomando en cuenta el crecimiento del PIB, fueron los años desde 1961 hasta 1981 (Banco Mundial, 2022). México tuvo crecimientos superiores a 4% en todo ese periodo.

En 1980, el valor de las exportaciones petroleras abarcaba 72% de las ventas, y para 1981 era 76%, subiendo en 1982 a 82%, así mismo en esos años, los impuestos petroleros representaban 25%, 26% y 44% de los ingresos ordinarios del gobierno federal, respectivamente, de acuerdo con (Colmenares, 2008). No obstante, aunque la dependencia petrolera hoy en día no representa el mismo nivel, sí tiene impacto en el desarrollo de las finanzas públicas, pues como podemos ver, en 2021 los ingresos estimados fueron de 6,295.7 MMDP⁴ (Congreso General de los Estados Unidos Mexicanos, 2020) de los cuales, el ingreso del sector público presupuestario fue 5,960.9 MMDP y los petroleros, 1,156.5 MMDP, por lo tanto, 19.40% de los ingresos provinieron aún del sector petrolero (SHCP, 2022). En ese año, el promedio del precio diario de la Mezcla Mexicana fue \$64.66, con una desviación estándar de \$6.98.⁵

(Ruiz-Porras & Anguiano-Pita, 2021) hacen un estudio entre la relación de los precios de petróleo y la actividad económica en México y establecen en sus conclusiones que el petróleo tiene implicaciones en la política económica para México. Por tal razón, el conocimiento a futuro de los precios de la Mezcla Mexicana de petróleo podría ayudar a tener finanzas públicas más estables, debido al nivel que este recurso representa en los ingresos totales de la Federación (Congreso General de los Estados Unidos Mexicanos, 2020). Si tuviéramos un pronóstico que se acercara al precio real del petróleo, o en su defecto, un pronóstico que nos permitiera ver la dirección del precio para la siguiente observación, podría ayudarnos a tener ingresos presupuestarios del sector público más estables, con ayuda del mercado de derivados. Por tanto, un objetivo de este trabajo es ayudar al lector a usar los modelos planteados como herramienta para la toma de decisiones. Si bien, esto es un grano de arena para el basto conocimiento, podría ser la inspiración para algunos lectores.

A manera de reflexión, citando al (Instituto Belisario Domínguez, 2017), su análisis consideró que las coberturas petroleras son una medida acertada para proteger las

⁴ Miles de Millones de pesos.

⁵ Elaboración propia con Datos de Banco de México. Los datos se extraen en la siguiente página:
<https://www.banxico.org.mx/apps/gc/precios-spot-del-petroleo-gra.html>

fluctuaciones de los ingresos petroleros, sin bien, comentan que los ingresos han superado sus costos, también es cierto que del 2008 al 2018 (años en los que se ha contratado una cobertura petrolera) solo en 3 años se han ejercido (2009, 2015, 2016) y en los otros años solo se ha pagado la prima por la contratación de las opciones tipo *put*⁶, lo cual representa un costo al no ejercerse.

⁶Significa opción de venta y ésta otorga al comprador el derecho, pero no la obligación de vender en el futuro un activo a un precio determinado previamente.

2. Pregunta de investigación e hipótesis

2.1 Pregunta de Investigación

Hace décadas, la programación de algoritmos usualmente estaba limitada por los requerimientos computacionales, así como por las técnicas empleadas, sin embargo, a lo largo de los años, los avances computacionales han hecho que los algoritmos de *Machine Learning* (en adelante ML), como las redes neuronales, hayan aumentado su uso debido al poder de procesamiento en las computadoras modernas, tal como lo plantea (Schmidt, Marques, Botti, & Marques, 2019).

Aproximarnos al precio de lo que pasará al menos una observación después sería de gran ayuda para el tomador de decisiones a efectos de contratar las coberturas necesarias, es por ello que debemos preguntarnos ¿Podrán tener las redes neuronales artificiales un desempeño superior al de un modelo estadístico de regresión lineal múltiple?

En este trabajo se emplearán datos desde enero de 1996 hasta junio del 2022, con periodicidad mensual dan un total de 318 observaciones por cada variable, éstas serán transformados en su logaritmo natural y diferenciadas en un periodo. Las variables WTI (siglas de *West Texas Intermediate*), BRENT⁷, DJI (siglas de *Dow Jones Industrial Average*), XAU (oro al contado en dólares estadounidenses⁸) y OilC1 (Se refiere al contrato 1 de futuros del *Cushing, OK Crude Oil*⁹) se usarán para pronosticar los precios de la Mezcla Mexicana de Petróleo y se comparará con un modelo base.

⁷ Petróleo de referencia en el continente europeo.

⁸ En inglés se conoce como *Gold Spot USD Dollar* y por convención dichas operaciones liquidan en 48 horas.

⁹ En dólares por barril.

2.2 Hipótesis

En la publicación de (Hill, Marquez, Marcus, & Remus, 1994) nos proponen a las redes neuronales artificiales como mejor herramienta para hacer pronósticos que un modelo estadístico, por ello, suponemos que esta técnica debería ser mejor que un modelo econométrico de regresión lineal múltiple tomado como base, por tanto, suponemos que las redes neuronales artificiales son mejores que una regresión lineal múltiple para pronosticar los precios de la mezcla mexicana de petróleo.

Para efectos de la evaluación del desempeño de los modelos, emplearemos como medida al RMSE¹⁰.

¹⁰ *Root mean-square error.*

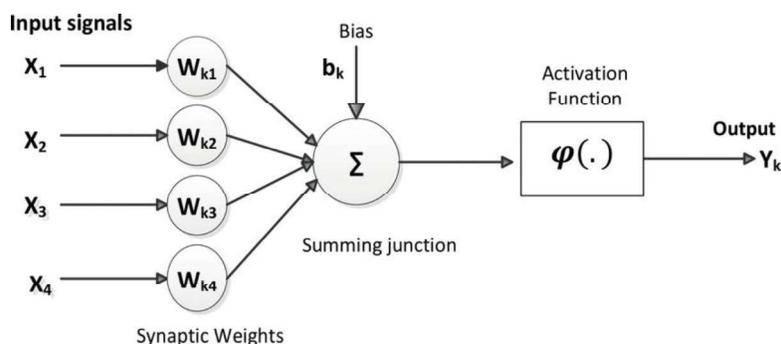
3. Revisión de literatura y marco teórico

3.1 Revisión de Literatura

Si visualizamos el nivel de procesamiento de las computadoras hoy en día, se podría pensar que el diseño de las redes neuronales es moderno, no obstante, los comienzos de las redes neuronales se remontan a la década de 1940 como lo señalan (McCulloch & Pitts, 1943), además el término actual *Deep Learning*¹¹ se acuñó en 1952 de acuerdo con (Huang, Chai, & Cho, 2020).

Una red neuronal artificial es una técnica en la que se empezó su diseño desde (McCulloch & Pitts, 1943) derivada de la inspiración de una red neuronal biológica (Véase ilustración 1). Hoy en día, las redes neuronales artificiales cuentan con una base matemática más sólida.

Ilustración 1 Celda de una Red Neuronal



Fuente: (El-Shahat, 2018).

En 1943 (McCulloch & Pitts, 1943) desarrollaron las primeras ideas de cómo las neuronas pueden funcionar dada la teoría neuropsicológica del sistema nervioso, posteriormente, Donald Hebb, en su escrito *Organization of behavior*, introduce los conceptos de "Sinapsis de Hebb", "ensamblaje celular Hebbiano" y "fase secuencial", de acuerdo con lo publicado por (McCulloch & Pitts, 1943).

(Brown, 2020) comenta que entre 1932 y 1945, se desarrollan las ideas sobre la estructura de la organización sináptica, que es un precedente de las redes neuronales. (McCarthy, Minsky, Rochester, & Shannon, 2006) señalan que Nathaniel Rochester llevó a cabo los primeros esfuerzos en IBM en los años 1950s y que el *Dartmouth Summer Research Project* se encargó de darle publicidad a los temas de inteligencia artificial y redes neuronales.

De acuerdo con (Peters, 2018), John von Neumann, en su escrito de 1958 *The Computer & The Brain* propone imitar las funciones de una neurona simple; este trabajo daría pauta al interés e imaginación de otros investigadores para continuar con el desarrollo de esta área en posteriores generaciones. Según (Peters, 2018), cuando Von Neumann agonizaba, realizó una crítica sobre la analogía entre el cerebro y las computadoras; más

¹¹ Traducción: Aprendizaje profundo.

de 60 años después, podemos ver más cerca la unión entre la tecnología computacional y el cerebro (BBC News Mundo, 2020).

Según (Widrow & Lehr, 1993), en 1959, Bernard Widrow y Marcian Hoff aplicaron los primeros modelos de redes neuronales a un problema de la vida real. Dichos modelos se llamaron ADALINE y MADALINE. Lo que conocemos como perceptrón fue criticado y finalmente descrito como limitado según (Minsky & Papert, 1970), sin embargo, el diseño del perceptrón que se parece más a una red neuronal moderna proviene del psicólogo Frank Rosenblatt y fue una red entrenada en la Universidad de Cornell en 1957. Dicho perceptrón tenía solo una capa que ajustaba los pesos y los límites de acuerdo con (Hardesty, 2017).

En 1982, John Hopfield presenta en la *National Academy of Science* un artículo en donde habla sobre el enfoque que pueden tomar las redes neuronales y la creación de distintos dispositivos (Hopfield, 1982). En 1985, el *American Institute of Physics* propició a la divulgación de las redes neuronales en una conferencia anual a nivel nacional, no obstante, a nivel internacional se abre en 1987 la primera conferencia de física IEEE¹² acaecida en San Diego. En ella participó la INNS¹³. Y así es como llegamos a uno de nuestro modelo de estudio. Dicho modelo es el LSTM (*Long Short-Term Memory* por sus siglas en inglés) *Neural Networks*¹⁴. De acuerdo con (Hochreiter & Schmidhuber, 1997) la estructura fue diseñada en 1997 debido a los tiempos de respuesta que tomaban las RNN¹⁵ por *back-propagation*¹⁶ y el costo computacional que estas acarreaban, además de que el algoritmo aprende más fácilmente, puede resolver tareas y encontrar patrones cuando hay una mayor cantidad de rezagos, es decir, almacenando conexiones de las entradas recientes de información en forma de activaciones (*Short-Term Memory*¹⁷ en contraposición a *Long-Term Memory*¹⁸).

Por su parte (Schuster & Paliwal, 1997) en 1997, presentaron en su trabajo *Bidirectional Recurrent Neural Networks*, una extensión a la estructura de red neuronal recurrentes en redes neuronales bidireccionales recurrentes (BRNN por sus siglas en inglés)¹⁹ en la que concluye que se puede entrenar la red en ambas direcciones simultáneamente y que usualmente ofrece mejores resultados que otras estructuras de redes neuronales artificiales (ANN por sus siglas en inglés.)

¹² *Institute of Electrical and Electronics Engineers. International Conference on Neural Networks.*

¹³ *International Neural Network Society.*

¹⁴ Redes Neuronales con memoria largo-corto plazo en español.

¹⁵ *Recurrent Neural Network* (en español: Redes Neuronales Recurrentes).

¹⁶ A diferencia del *feed-forward*, el *back-propagation* no tiene un flujo en una sola dirección que va de las capas de entrada a las de la salida, sino que, una vez obtenido el error, vuelve a la capa anterior (una propagación hacia atrás), es decir, emplea un ciclo propagación – adaptación en dos fases.

¹⁷ Memoria de corto plazo.

¹⁸ Memoria de largo plazo.

¹⁹ Cabe señalar que en el presente trabajo se usa la nomenclatura BILSTM.

3.2 Marco Teórico

Primeramente, necesitamos saber qué variables vamos a usar para pronosticar el precio de la Mezcla Mexicana, esto último lo lograremos mediante una regresión lineal múltiple con la que analizaremos si las variables propuestas son significativas. Después, recrearemos un modelo SARIMA²⁰ (por sus siglas en inglés) y un modelo ARDL²¹ (por sus siglas en inglés) para visualizar qué variables rezagadas son importantes y así darnos una idea de cuántos rezagos podemos incorporar al modelo y cuáles variables a la red neuronal; adicionalmente, haremos una valoración de las variables significativas del modelo de ML usando el algoritmo F Test de Matlab, y ver si los resultados son similares a los de los métodos econométricos mencionados.

Cabe señalar que los métodos de econometría usados aquí, solo se presentarán para visualizar las variables que pudieran impactar en el modelo para el pronóstico de la red neuronal, por lo tanto, no ahondaremos en las pruebas que debe pasar la regresión, aunque sí compararemos las variables significativas de econometría con los resultados que obtengamos del F Test. Hay que recordar que ML se nutre de la relación con los datos, por tanto, es otro su enfoque.

Las variables que se seleccionaron en el modelo son el WTI (siglas de *West Texas Intermediate*), BRENT (petróleo de referencia en el continente europeo), DJI (siglas de *Dow Jones Industrial Average*), XAU (oro al contado en dólares estadounidenses) y OILC1²² (Se refiere al contrato 1 de futuros del *Cushing, OK Crude Oil*) para pronosticar la mezcla mexicana fueron tomadas en cuenta porque son variables que se utilizaron en otras investigaciones²³ como (Chen, He, & Tso, 2017) o (Gupta & Nigam, 2020) e inclusive (Daneshvar, Ebrahimi, Salahi, Rahmaty, & Homayounfar, 2022) y (Villada, Arroyave, & Villada, 2014), además de (Sivaprakash & Manu, 2021) y (Yao, Pu, & Qiu, 2021), pues resultaron ser significativas para estimar otros precios de petróleo de otros países.

Las conexiones de la redes neuronales tienen una naturaleza no lineal y están asociadas con un peso (W). Mediante un proceso iterativo y adaptativo, a lo que llamaremos aprendizaje, el algoritmo irá modificando los pesos (al principio son aleatorios), de acuerdo con las entradas suministradas, de tal forma que encuentre el conjunto de datos y conexiones que coincida lo más posible a los resultados requeridos, después de que se proporcione una salida (Pérez Ramirez & Fernández Castaño, 2007).

Para conseguir lo anterior, es necesario conocer la diferencia entre los valores predichos y el valor correcto, esto se encuentra relacionado con lo que es la “función de costos”. El algoritmo cambia los pesos de tal forma que obtengamos la función de costos mínima posible (Han, Kim, & Youn, 2018).

²⁰ Seasonal *Autoregressive Integrated Moving Average* (un modelo que incluye como variables explicativas a medias móviles y autorregresivos de la misma variable explicada además del componente estacional).

²¹ *Autoregressive Distributed Lag* (en econometría, es un modelo que incluye rezagos de variables independientes diferentes a la variable dependiente).

²² Definidas en la siguiente sección del trabajo.

²³ O han tomado variables similares homologados a su país de origen.

Se separa la base de datos en entrenamiento y prueba porque los datos de entrenamiento se usan para construir el modelo, mientras que los datos de prueba sirven para evaluar el modelo según los resultados que ofrezca el modelo entrenado. Existen muchos tipos de separación de datos como lo podemos ver en las minutas de la conferencia de (Muraina, 2022).

Los pesos son actualizados durante el entrenamiento a una tasa denominada *step size*²⁴ o *learning rate*²⁵, que usualmente va de 0.0 a 1.0, y esto es considerado un hiperparámetro²⁶. Hay que entender que, entre menor es el *learning rate*, se va a requerir de un mayor número de *epochs*²⁷ dado que los cambios en los pesos van a ser muy pequeños y por tanto requerirá mayor tiempo de entrenamiento. El número de *epochs* es un hiperparámetro del gradiente descendiente²⁸ que controla el número de veces que pasa por completo el conjunto de datos de entrenamiento, pero no hay que confundir con el *batch size*²⁹ que controla las muestras del entrenamiento a trabajar antes de que se actualicen los parámetros internos del modelo.

De acuerdo con (Manjarrez, 2014), las capas son la estructura con la cual está diseñada una red neuronal y cada capa contiene un número determinado de neuronas, cómo aparece en la ilustración 1 anterior. Se pueden distinguir tres tipos de capas. En primer lugar, la capa de entrada la cual recibe los datos que están en contacto con el entorno. En segundo lugar, la capa oculta que no tiene conexión con el entorno y puede ser precedida por otras capas ocultas o por la capa de entrada. En tercer lugar, la capa de salida que proporciona la solución de la red neuronal. Véase la ilustración 2.

²⁴ Se mantienen los conceptos en inglés porque la mayoría de la literatura está en ese idioma.

²⁵ *Step size* se refiere a cuánto se mueve el factor que cambia los pesos. $W = W + (-step_size) * dJdW$ donde $dJdW$ es el *Gradient of Loss*. Por otro lado, el *learning rate* es el hiperparámetro que controla cuánto cambia el modelo en respuesta a un error estimado cada vez que los pesos (w) son actualizados.

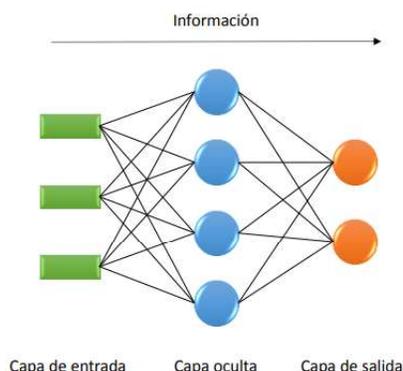
²⁶ Son los valores que se utilizan para entrenar a los modelos, se distingue de los parámetros porque los parámetros son los valores que se obtienen durante el proceso de entrenamiento.

²⁷ Época en español, es otro hiperparámetro que consiste en el número de veces que pasan los datos de entrenamiento por el algoritmo de *machine learning*, usualmente los datos están agrupados en *batches*

²⁸ Se entiende como el algoritmo de aprendizaje iterativo que usa los datos del entrenamiento para actualizar el modelo. El SGD se conoce como *Stochastic Gradient Descent* y es un algoritmo de optimización usado en el entrenamiento de los modelos de ML.

²⁹ Tamaño del lote en español, consiste en el número de muestras de entrenamiento utilizadas en cada iteración. Cuando el *batch size* es igual que el total de datos, hacer la iteración y el *epoch* son valores equivalentes.

Ilustración 2 Capas de una Red Neuronal

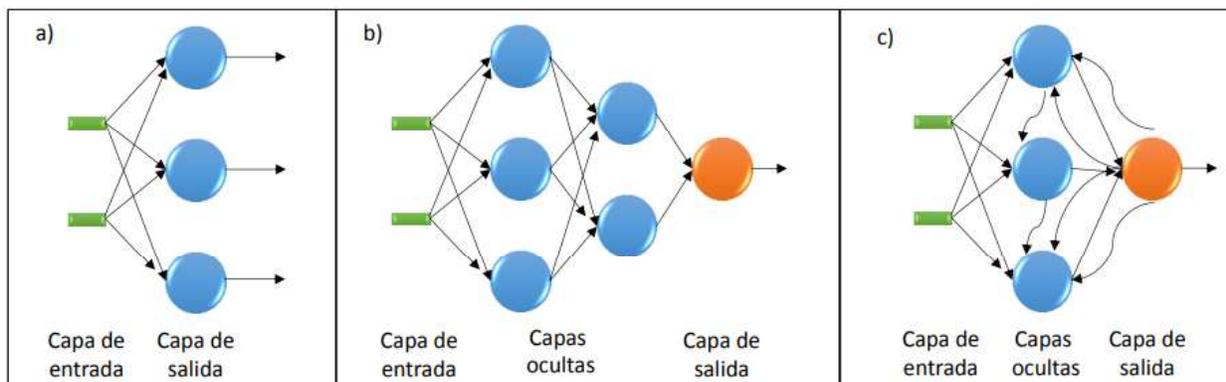


Fuente: (Manjarrez, 2014).

Una red monocapa solo proyecta las entradas de una capa de neuronas con la salida donde se realizan los cálculos, mientras que una red multicapa contiene un número determinado de capas intermedias a la capa de entrada y de salida (capas ocultas). Esto lo podemos ver en (Manjarrez, 2014).

De acuerdo con (Manjarrez, 2014), una red neuronal no recurrente, a diferencia de una recurrente, se puede distinguir porque la primera tiene la propagación de señales solo en un sentido y no tienen un sistema de realimentación. Véase la ilustración 3.

Ilustración 3 a) Red Neuronal monocapa b) Red Neuronal multicapa no recurrente con propagación hacia adelante c) Red Neuronal multicapa recurrente



Fuente: (Manjarrez, 2014).

Según (Pérez Ramirez & Fernández Castaño, 2007), la red más utilizada es el perceptrón multicapa (*Multi-Layer Perceptron*), usualmente una red de 3 capas (entrada, oculta y salida), que cuenta con funciones de activación como función de transferencia en la capa oculta y dos funciones para la capa de salida que pueden ser lineales o sigmoideas, dependiendo del tipo de salida que se requiera.

(Zaremba, Sutskever, & Vinyals, 2015) describen a una red neuronal recurrente (RNN de ahora en adelante), como un modelo de secuencia neuronal que puede ser descrito usando transiciones determinísticas de estados previos a ocultos. La transición del estado determinista de la función aparece en (1), en donde los subíndices denotan el

$timestep^{30}$ t y los superíndices denotan las capas, todos los estados tienen una dimensión n , además $h_t^l \in \mathbb{R}^n$ es el estado oculto de la capa l , i.e. l es el número de capas en la red neuronal:

$$RNN: h_t^{l-1}, h_{t-1}^l \rightarrow h_t^l \quad (1)$$

La RNN, está dada por la función (2), en la que $f \in \{sigm|tanh\}$ es la función de activación³¹:

$$h_t^l = f(T_{n,n}h_t^{l-1} + T_{n,n}h_{t-1}^l), \text{ donde } f \in \{sigm|tanh\} \quad (2)$$

En las redes neuronales artificiales, la capa de entrada (*input layer*) es la capa donde entra la información y es transferida a la segunda capa (*hidden layer*³²), después de haberse procesado la información pasa a la capa de salida (*output layer*), así lo explica (Han, Kim, & Youn, 2018). Se puede ver un ejemplo en la ilustración 2 anterior, la cual contiene 3 capas, mientras que en la ilustración 1 anterior se pueden ver los pesos (W), la función de activación y el sesgo.

(Hochreiter & Schmidhuber, 1997), en un artículo publicado por el MIT, señalan que el algoritmo LSTM tiene la capacidad de superar retrasos más largos en las variables de entrada, así como la generalización y el manejo de ruido de las series y valores continuos, de igual forma, tiene la capacidad de aprender más rápido aunque haya datos de entrenamiento con poco rezago y tiene bastante sensibilidad con respecto al *time-step*³³, pero también tiene limitaciones, como que cada celda de memoria necesita dos unidades adicionales de puertas de entrada y de salida, de igual forma, en comparación con las redes neuronales recurrentes, el número de pesos³⁴ se incrementa exponencialmente, así mismo, el algoritmo sufre la incapacidad de contar con precisión los patrones de *time-step* discretos si las entradas tienen mucho rezago una de la otra, pues no ayudan a reducir el error esperado. Para reducir el error, en el trabajo de (Hochreiter & Schmidhuber, 1997) se hicieron varios experimentos, se generaron pesos que tuvieron un rango de [-0.2 a 0.2] y [-0.1 a 0.1], las secuencias de entrenamiento fueron generadas aleatoriamente, además, cada *time-step* discreto de cada secuencia de entrada envuelve tres pasos: 1) usar la entrada recurrente y establecer las unidades de entrada; 2) calcular las funciones de activación de las capas ocultas y 3) calcular las activaciones de las unidades de salida. La arquitectura de la red LSTM tuvo un enfoque sistémico, es decir, la capa oculta empieza con una neurona, después aumenta a dos y así sucesivamente. Dicha red, tiene celdas neuronales en las que contienen puertas de entrada y salida, por tanto, el algoritmo aprende cuándo abrir y cerrar los accesos a los flujos de información. Esto ayuda a que en su salida pueda ser protegida de la perturbación de otras unidades en caso de contenido irrelevante.

³⁰ La muestra que se va a usar como variable de entrada con respecto a que tan larga es en el tiempo. En español, pasos del tiempo.

³¹ También se le conoce como función de transferencia.

³² Las capas ocultas son conocidas como *black box* por la compleja interpretación de cómo la Red Neuronal Recurrente (RNN por sus siglas en inglés) deriva de un resultado en particular.

³³ Cuántas observaciones quieres mirar atrás o cuántos puntos de datos del pasado se quieren considerar añadidos al dato recurrente para hacer una predicción.

³⁴ *Weight* en inglés.

Aunque los algoritmos de ML, en este caso Redes Neuronales han tenido mucha fama en los últimos años, existen modelos de ML de regresión que suelen ser complementarios a la hora de hacer pronósticos, en este caso los modelos ARIMA. (Shah & Kiruthiga, 2020) examinan series de tiempo no lineales de los precios de petróleo usando estos modelos ARIMA, para ello usan el WTI como variable dependiente en su logaritmo natural y como variables explicativas, la misma variable WTI rezagada, el Brent y el Dubai/Oman *oil benchmarks*. En su trabajo, los autores descomponen la serie en componentes y aplican 80% de la información en entrenamiento y 20% en pruebas, para un periodo que abarca desde julio 1987 a marzo 2020. Para ellos, la diferenciación óptima se encuentra en 1 para convertirla en estacional y al analizarlo, terminan con un modelo en donde $AR = 0$, $I = 1$ y $MA = 4$. Las limitaciones de este tipo de modelos, y los mismos autores lo expresan, se encuentra en los años en donde las series tienen altas volatilidades, pues capturar este tipo de cambios lo hace complicado.

Además de los modelos ARIMA, existen otras alternativas. En el trabajo presentado por (Yao, Pu, & Qiu, 2021), los autores usan el algoritmo LSTM, pero con dos capas ocultas, con datos de la OPEP desde enero de 2003 hasta septiembre del 2021, teniendo un total de 4,756 datos diarios (90% entrenamiento), el optimizador es ADAM³⁵ y usan *TensorFlow*³⁶. Las dos capas tienen una densidad de 64 neuronas y el RMSE³⁷ y MAE³⁸ son usados para evaluar los modelos. Los *time-steps* utilizados son 10, 20 y 30; con ellos se alcanzó un RMSE de 3.741, 4.419 y 3.928 respectivamente, mientras que *Prophet*, de Facebook, alcanzó un RMSE de 3.212.

El optimizador ADAM es un algoritmo basado en gradientes de primer orden de una función objetivo estocástica. El método tiene pocos requerimientos computacionales, éste, calcula el *adaptive learning rate*³⁹ para diferentes parámetros de las estimaciones del primer y segundo momento de los gradientes. Así lo explica (Kingma & Ba, 2015).

Por otro lado, en la India desarrollaron un modelo de pronóstico y lo presentaron en la *Asian Journal of Management* para los precios del *Indian Crude Oil* a través de una Red Neuronal con mecanismo de *back-propagation* y *forward-propagation* (Sivaprakash & Manu, 2021). Ellos usaran como variables independientes la tasa de inflación del país, así como su tipo de cambio USD/INR, exportaciones e importaciones de crudo en la India, el Nifty Index⁴⁰, así como la producción de petróleo, y los precios de las acciones de los 6 mayores productores de petróleo de dicho país. En este caso, ellos usan el método de regresión múltiple para identificar las variables de mayor impacto. A pesar de que el *feed-forward* tuvo un desempeño en el entrenamiento superior, la que tuvo el mejor desempeño en las pruebas fue el *back-propagation*. Añadido a lo anterior, los autores recomiendan usar algoritmos de RNN, Redes Neuronales Convolucionales (CNN por las siglas en inglés) y de LSTM. Finalmente, los modelos arrojaron un RMSE para las pruebas de *back-propagation* de 2.35 y para el *feed-forward* de 3.93.

³⁵ Es un acrónimo de su versión inglesa *adaptive moment estimation*.

³⁶ De Python.

³⁷ *Root Mean Squared Error*.

³⁸ *Mean Absolute Error*.

³⁹ En español: tasa de aprendizaje adaptativo.

⁴⁰ Índice de precios y cotizaciones de la India.

(Kulkarni & Haidar, 2009) emplea la metodología de una red neuronal *feed-forward* y el algoritmo de *back-propagation*, con una y dos capas ocultas, el cual busca el pronóstico de tres días hacia adelante del WTI, tomando como variables independientes las mismas variables WTI rezagadas, así como el precio de los futuros de los contratos, y toma en cuenta los precios al cierre de septiembre de 1996 hasta agosto de 2007. Tenemos un total de 2,705 datos de cada serie con una periodicidad diaria, de los cuales 90% se usa para entrenamiento y 10% para pruebas. El objetivo es el pronóstico para la siguiente observación. La base de datos pasó por un proceso de estandarización max-min; de igual forma, se usa el RMSE como en las demás publicaciones. Como función de activación se usa una función sigmoide simétrica. Para su metodología, se corren de una a 10 neuronas en el modelo y cada red se corre al menos tres veces antes de escoger otro *set* de pesos⁴¹ además de recorrerlo de uno a 20 rezagos en las variables para la capa de entrada. La última observación para destacar es el uso del método *multi-step*⁴² pero de acuerdo con sus conclusiones, tiene cierta validez solamente hasta dos observaciones adelante, por la precisión del modelo a dos observaciones. Por otro lado, lo mejor para el *benchmark* fueron 7 rezagos para la entrada con una estructura de 3 capas y 8 neuronas en las capas ocultas. El enfoque para crear el *benchmark* se basa en los valores rezagados del precio del crudo *spot*⁴³ preprocesado, posteriormente, se añaden esos valores preprocesados a los futuros de 1, 2, 3 y 4 meses.

Aunque se puede hacer un algoritmo de ML simplemente usando como variables de entrada a las variables rezagadas de la misma variable que queremos predecir, (Gupta & Nigam, 2020) señalan que es conveniente añadir indicadores fundamentales para el pronóstico de los precios. Esos autores estuvieron trabajando en un algoritmo de redes neuronales con *back-propagation*, en este algoritmo usaron los precios de cierre en un periodo de 5 años con 11 meses, Su recolección fue de 1,500 datos diarios, dichos datos fueron normalizados, además se usó el criterio RMSE como criterio de evaluación y se separó 70% de los datos para entrenamiento. Cabe señalar que los modelos LSTM tienen una realimentación de datos, al igual que las redes neuronales recurrentes por ser de una arquitectura similar, por lo que son estructuras con memoria, tal como lo comenta (Manjarrez, 2014) pero en específico, como lo expresan (Hochreiter & Schmidhuber, 1997), la red del tipo LSTM tiene la capacidad de superar retrasos más largos en las variables de entrada.

(Gupta & Nigam, 2020) destacan que el modelo con menor RMSE se alcanzó con 3 rezagos y aunque los autores establecen que la predicción del precio exacto es muy complicada, las redes neuronales son efectivas para pronósticos de precios de corto plazo y generar estrategias de inversión.

Del modelo se desprende que se pueden usar como variables independientes otras cotizaciones del precio del petróleo distintas a la que se quiere pronosticar y no solo la misma variable rezagada en el tiempo de acuerdo a las recomendaciones de (Gupta & Nigam, 2020), además de existir el apoyo de índices bursátiles, cotizaciones de

⁴¹ *Weights* (pesos).

⁴² No solo pronostica 1 observación, sino que usa la variable pronosticada como *input* para volver a pronosticar la observación subsecuente.

⁴³ Se refiere al Tipo Valor que en este caso son 48 horas, que es lo mismo que decir que su liquidación es "t+2".

empresas petroleras y cotizaciones de divisas, como el *U.S. Dollar Index* (DXY⁴⁴), en la publicación realizada por (Villada, Arroyave, & Villada, 2014).

Por tanto, las variables que se pueden utilizar en redes neuronales para explicar el comportamiento del precio de barril de petróleo son diversas, no obstante, se pueden usar distintos rezagos y otras variables como índices bursátiles, a la manera de (Gupta & Pandey, 2018) quienes emplean datos semanales para pronosticar el rendimiento del precio del barril de petróleo. Ellos usan al rendimiento del índice NIFTY, así como el precio del oro. En su publicación, usan una profundidad mayor a dos capas ocultas, pues en este caso su estructura se basó en 4 capas de LSTM con 32 neuronas en cada una, que conectan a una capa tradicional con densidad 10 y a una capa final con densidad uno antes de la salida de los datos; cabe señalar que el uso de 10 rezagos y solamente 100 *epochs* fueron suficientes para encontrar el mejor modelo, además, una de las conclusiones del modelo fue que al aumentar el número de capas, realmente no mejora la exactitud⁴⁵ del modelo; por otro lado, comentaron que en sus próximas publicaciones considerarían las condiciones del mercado y el ambiente político dentro de los modelos.

Los modelos de mayor profundidad en las redes neuronales, es decir, de más de una capa, han demostrado que han dado mejores resultados, según lo publican (Daneshvar, Ebrahimi, Salahi, Rahmaty, & Homayounfar, 2022). Los autores hacen un pronóstico para el marcador BRENT, usando datos diarios de enero de 2015 a marzo de 2022, aunque el aspecto a destacar es la comparación que hacen de un modelo LSTM tradicional en contraposición al *Bidirectional Long-Short Term Memory* (BI-LSTM⁴⁶ de ahora en adelante), el cual consta de dos capas ocultas, en este caso usan diferentes métodos de optimización, entre ellos está el ADAM⁴⁷, RMSProp⁴⁸ y SGDM⁴⁹. Entre las conclusiones del estudio se encuentran el hallazgo de la correlación que tiene el BRENT con el precio del oro, así como con el dólar canadiense. De igual forma, se comenta que el rendimiento de estos algoritmos depende del tipo y calidad de entrenamiento, aunque en esta situación, el tipo de optimización SGDM fue superior a las de los otros métodos. En este caso, añadir otra capa mejora la exactitud⁵⁰ del modelo. Por último, un rasgo a destacar es que en esta publicación se tomó en cuenta la caída de los precios de petróleo dada la sobreoferta por el COVID-19.

Al analizar a (Gupta & Pandey, 2018) que usan el algoritmo de LSTM y a (Daneshvar, Ebrahimi, Salahi, Rahmaty, & Homayounfar, 2022) que emplean el BI-LSTM, nos percatamos que las arquitecturas LSTM y BI-LSTM son similares, pero no iguales. En la ilustración 4 aparecen ambos modelos de redes neuronales, pero el algoritmo LSTM es

⁴⁴ Es el índice de referencia de diferentes divisas con respecto al dólar estadounidense.

⁴⁵ El término original utilizado en la publicación de (Gupta & Pandey, 2018) fue descrita por la palabra *accuracy*. Se hace este paréntesis porque en ML es diferente *accuracy* y *precision*. Por tanto, se tradujo en este párrafo como exactitud en vez de precisión.

⁴⁶ Es un algoritmo de procesamiento de secuencias que consta de dos LSTM: uno que toma la entrada en *forward propagation* y otra en *backward propagation*.

⁴⁷ Es un algoritmo de optimización diseñado específicamente para entrenar redes neuronales profundas (*Deep neural network*).

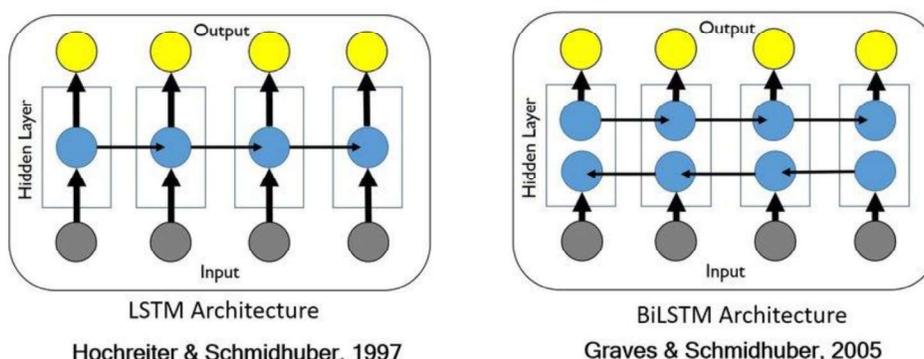
⁴⁸ Método de optimización *Root Mean Squared Propagation*.

⁴⁹ *Stochastic Gradient Descent with Momentum*.

⁵⁰ Mismo caso que el anterior, se usa la palabra *accuracy* en la publicación respectiva.

unidireccional mientras que, si al LSTM se le añade otra capa, se puede revertir la dirección del flujo de la información y se convierte en un BI-LSTM. En el modelo BI-LSTM de la ilustración 4 se puede ver a un *backward layer* en contraposición al *forward layer*⁵¹.

Ilustración 4 Arquitectura de una Red Neuronal LSTM y BI-LSTM



Fuente: (Mohand & Gaitonde, 2018).

En 2017, (Abadi, Isard, & Murray, 2017) ofrecieron una introducción para usar TensorFlow, una herramienta de ML empleada en Python. Dicha herramienta fue aprovechada por (Chen, He, & Tso, 2017) cuando desarrollaron otro modelo de pronósticos del WTI (Modelos híbridos⁵²), en los que solo se usan las variables rezagadas en ciertos periodos. Los modelos híbridos han comenzado a tomar fuerza desde entonces. En la publicación, obtienen un pronóstico siguiendo la expresión $y_t = w_{lm}\hat{r}_{lm} + w_{nlm}\hat{r}_{nlm}$ en el que y_t es el precio del WTI, w representa la contribución (o peso) que se les da a las estimaciones r , en este caso \hat{r}_{lm} son las estimaciones con modelos lineales como ARIMA y Caminata Aleatoria, en contraste, \hat{r}_{nlm} proveniente de modelos no lineales, en este caso representado por el algoritmo *Deep Belief Network* (DBF de ahora en adelante) y LSTM. En dicha publicación, usan una w de 0.5 en cada una. Se usan datos desde julio del 2007 a febrero del 2017, teniendo un total de 2,409 observaciones diarias, de éstas, 70% es de entrenamiento. Los datos se transforman a su logaritmo natural y además se estandarizan. Una de sus conclusiones fue que el modelo DBF es muy funcional para pronosticar los precios de petróleo y una de sus características principales, a diferencia de los demás, es que es un algoritmo multicapa y, aunque hay conexión entre las capas, no existe interconexión entre las neuronas de la misma capa. Así mismo, el DBN se diseñó con 4 capas ocultas con una densidad de 100 neuronas⁵³ en cada una, aunado a una función de activación sigmoide y un Learning Rate de 0.01, mientras que en el modelo LSTM se usaron 2 capas ocultas con 4 neuronas en cada una, más un algoritmo de optimización ADAM y el MSE como función de pérdida.

⁵¹ En el *backward layer* el flujo es de atrás hacia adelante, mientras que en el *forward layer* el flujo va hacia adelante.

⁵² Los modelos híbridos combinan diferentes algoritmos, procesos o procedimientos de un dominio o área de conocimiento similar o diferente. Tienen por objeto complementarse mutuamente para resolver un problema.

⁵³ O Celdas.

Otra de las conclusiones fue que el LSTM producía pronósticos con rangos de fluctuación grandes, por lo que los autores creen que no sería apropiado cuando ocurran cambios muy rápidos; además, el mercado tiene corta memoria y ese comportamiento es predominante y, al igual que otros autores, opinan que hacer más complejos los modelos con más capas ocultas o más neuronas en cada capa, no necesariamente resulta en una mejoría.

Los modelos híbridos en los que se usan ARIMA en conjunto con el LSTM como algoritmo de ML, no son exclusivos del mercado petrolero, sino que se han usado para pronosticar otras variables como el S&P500. Así lo refieren (Kulshreshtha & Vijayalakshmi, 2020), para ello emplean datos del S&P500 desde 2015 a 2019 para entrenamiento y un trimestre del 2020 para pruebas; los autores usan las variables del precio de apertura, el nivel más alto alcanzado en ese día y lo más bajo alcanzado, así como el cierre y su volumen de operación, y emplean el RMSE y el MAE para evaluar el modelo. Aunque no mencionan si estandarizan la serie, sí señalan que hacen una transformación logarítmica. En este caso, una de las conclusiones es que el modelo híbrido arroja mejores resultados que los modelos individuales. De los elementos a resaltar de esta publicación es que el modelo híbrido lo hacen a partir del ARIMA, es decir, usan los residuales del modelo ARIMA para entrenar el modelo LSTM y usan el modelo LSTM para mitigar el *vanishing gradient problem*⁵⁴, lo cual es común en las redes neuronales.

En las redes neuronales, a diferencia de un modelo econométrico, la multicolinealidad es una característica que puede ayudar al pronóstico. Así podemos observarlo con (Meneses-Bautista & Alvarado, 2017), quienes diseñan una red neuronal de *back-propagation* para pronosticar el tipo de cambio dólar-peso (USD/MXN de ahora en adelante). Cuando se usan los tipos de cambio del peso con respecto a otras divisas, tenerlos como variables independientes para explicar el USD/MXN es una decisión adecuada, aunque haya alta correlación entre las variables explicativas en el modelo. En este modelo emplean el tipo de cambio del peso con respecto al euro, yen y libra esterlina, además del uso del USD/MXN como variable de rezago para explicar el USD/MXN en el periodo $t+1$. El periodo de prueba corre desde enero de 2015 hasta marzo de 2017. Los autores separan la base de datos en 3: semestral, 1 año y 2 años, e independientemente de cuantas observaciones sean de periodicidad diaria, se dejan 30 objetos⁵⁵ para las pruebas. Posteriormente, se escalan los datos de 0 a 1, a excepción del yen japonés; posteriormente, el vector pasa por un proceso de estandarización min-max. Al final, los autores concluyen y apoyan la idea de no descartar ninguna de las 4 variables del tipo de cambio e inclusive, amplían la posibilidad de usar el tipo de cambio inverso, por tanto, aunque el uso de las variables no sea el adecuado desde el punto de vista econométrico debido a su multicolinealidad, sí pueden apoyar en el pronóstico cuando se emplea una red neuronal.

Entendemos a la predicción como la estimación de resultados que no se han visto aún, pero a diferencia del pronóstico, éste último, hace predicciones usando a las series de

⁵⁴ Cuando hay más capas en la red, el valor del producto de la derivada disminuye hasta que en algún momento la derivada parcial de la función de pérdida se acerca a un valor cercano a cero, y la derivada parcial desaparece.

⁵⁵ U observaciones.

tiempo como bases de datos, por tanto, el pronóstico usa la dimensión temporal para predecir, así como podemos verlo a manera de ejemplo en la siguiente publicación (Fotios et al, 2022) mientras que podemos ver un ejemplo de predicción en esta otra publicación (Chibani & Coudert, 2020).

Como lo hemos visto en los párrafos anteriores, las técnicas avanzadas de ML y las redes neuronales han demostrado ser apropiadas para las tareas de predicciones. Otra técnica de ML, llamada *Deep Learning* (DL⁵⁶ de ahora en adelante), posee ventajas para el entrenamiento de *big data*⁵⁷ debido a su fuerte capacidad de generalización y robusto poder, en comparación con otras técnicas convencionales como *Support Vector Machine* (SVM por sus siglas en inglés) y *K-Nearest Neighbors* (KNN por sus siglas en inglés). Las Finanzas y la Economía no han sido las primeras áreas de aplicación de DL, la técnica también ha sido abierta a otros campos. A manera de referencia, véase a (Li, Harfiya, Purwandari, & Lin, 2020) para la Medicina, o en Neurociencias véase (Lindsay, 2020), para la Física o Astronomía se puede consultar a (Baron, 2019).

De acuerdo con (Huang, Chai, & Cho, 2020), en la literatura financiera los algoritmos que están considerándose mayormente son: DL, RNN, CNN y LSTM.

Así como (Gupta & Pandey, 2018) y (Kulkarni & Haidar, 2009) publican el uso de los algoritmos usando una arquitectura con más de una capa oculta y presentan un mejor desempeño en el entrenamiento del algoritmo; en este trabajo usaremos una arquitectura de una y dos capas ocultas, además tomaremos en cuenta el algoritmo de LSTM por las bondades que tiene, por ejemplo, para solventar el problema del gradiente descendiente y la capacidad de poder “recordar” información del pasado, no obstante, no solo vamos a probar el modelo con una sola capa, sino que también probaremos con más profundidad e intentaremos dos capas.

⁵⁶ Aprendizaje profundo, aunque es mejor dejar el término en inglés para facilitar su uso, pues la mayor parte de la literatura actual se encuentra en inglés.

⁵⁷ Se refiere al gran volumen de datos y que muchas veces su procesamiento es a través de técnicas distintas a las convencionales.

4. Metodología - especificar la metodología

4.1 Selección de Datos

Se observaron las variables que tuvieron impacto en otras investigaciones descritas en el marco teórico y se escogieron las siguientes para el pronóstico de la Mezcla Mexicana:

1. La Mezcla Mexicana (dólares por barril) {MM de ahora en adelante}
2. *Cushing, OK WTI Spot Price* FOB (dólares por barril) {WTI de ahora en Adelante}
3. *Europe Brent Spot Price* FOB (dólares por barril) {BRENT de ahora en Adelante}
4. *Dow Jones Industrial Average* {DJI de ahora en Adelante}
5. *Gold Spot US Dollar* (XAU/USD) {XAU de ahora en Adelante}
6. *Crude Oil (Light-Sweet, Cushing, Oklahoma, Contract 1)* (dólares por barril) {OILC1 de ahora en Adelante}

En donde la variable dependiente es MM y las independientes son WTI, BRENT, DJI, XAU y OILC1, así como el MM con sus rezagos.

Una forma de visualizar las variables geopolíticas sería con la inclusión de variables que representen la volatilidad del sector energético como es el caso del *Energy Volatility Index* (EVL) el cual es publicado por CME⁵⁸ Group y puede verse la metodología en esta publicación (CME Group, 2022), no obstante, estos datos no nos indican si el precio va a subir o bajar, solamente estima mayor volatilidad. De igual forma las estadísticas proporcionadas por (IEA, 2022) en el *World Energy Outlook* nos provee información valiosa, no obstante, para obtener dicha información en ambos casos es necesario pagar por ella, además, la información que proporciona el *World Energy Outlook* no emite información mensual, el cual, es una periodicidad usada en este trabajo. En resumen, el contrato 1 del *Light-Sweet, Cushing, Oklahoma* al ser un contrato de derivados, será la variable más cercana que se usará como medida de volatilidad.

Los datos de la Mezcla Mexicana fueron extraídos de la base de datos del Banco de México, mientras que los datos del WTI⁵⁹ y BRENT⁶⁰, se extrajeron de Reuters, aunque se encuentran de igual forma en la página web de la *Energy Information Administration* (EIA por sus siglas en inglés) y lo mismo sucede con el contrato 1 de futuros del OK Crude Oil⁶¹. Por otro lado, los datos del DJI y XAU se descargaron de la página web www.investing.com

Los datos son mensuales y se tomó en cuenta la valuación al cierre del último día hábil del mes; para los días en los que no hubo cotización el último día hábil del mes, se tomó la valuación del cierre del día anterior ($t-1$) o el día anterior a ese ($t-2$) y así sucesivamente.

⁵⁸ Chicago Mercantile Exchange.

⁵⁹ Se puede encontrar la información del WTI aquí:

<https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=D>

⁶⁰ Se puede encontrar la información del BRENT aquí:

<https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBTE&f=D>

⁶¹ Aquí se puede encontrar el OILC1: <https://www.eia.gov/dnav/pet/hist/RCLC1D.htm>

Los datos se encuentran con una fecha de inicio de enero de 1996 y terminan en junio del 2022, por tanto, tenemos un total de 318 observaciones, de los cuales, usamos 22% como prueba y el resto para entrenamiento⁶². Todos los datos son transformados en su logaritmo natural. Hay que tomar en cuenta el número de rezagos por cada variable y su diferenciación en cada caso.

Empleando la estructura de (Meneses-Bautista & Alvarado, 2017), sea $f_t = f(x_t)$ el valor de una función en el momento t equivalente a y_t^{MM} , el precio de la Mezcla Mexicana, y x_t un vector cuyos componentes $(x_t^1, x_t^2, x_t^3, \dots, x_t^n)$ representan el conjunto de las n variables independientes que describen el fenómeno en el tiempo t .

Sea f_{t+1} el siguiente valor de función, obtenida de otra función g , dada las variables independientes x_t^n , como aparece en la ecuación 3, similar a (Meneses-Bautista & Alvarado, 2017):

$$f_{t+1} = g(x_t, F_t) = g(x_t^1, x_t^2, \dots, x_t^n, F_t) \quad (3)$$

Tomando en cuenta los datos del entrenamiento con respecto a las variables seleccionadas anteriormente para el vector x_t queda de la forma como aparece en la ecuación (4):

$$y_t^{MM} = (x_t^{MM}, x_t^{WTI}, x_t^{BRENT}, x_t^{DJI}, x_t^{XAU}, x_t^{OILC1}), \text{ con } x_t \in \mathbb{R} +^6 \quad (4)$$

Para el uso del modelo, con el fin de suavizar la serie, se elige una transformación por logaritmo natural, tal como aparece en la ecuación (5):

$$\ln(x_t) = \ln(x_t^{MM}, x_t^{WTI}, x_t^{BRENT}, x_t^{DJI}, x_t^{XAU}, x_t^{OILC1}) \quad (5)$$

$x_t^{MM} \in \mathbb{R} +$ es el valor de la mezcla mexicana

$x_t^{WTI} \in \mathbb{R} +$ es el valor del Cushing. OK WTI Spot Price FOB

$x_t^{DJI} \in \mathbb{R} +$ es el valor del Dow Jones Industrial Average

$x_t^{XAU} \in \mathbb{R} +$ es el valor del Gold Spot US Dollar (XAU/USD)

$x_t^{OILC1} \in \mathbb{R} +$ es el valor del Light-Sweet, Cushing, Oklahoma, Contract 1

⁶² En total son 317 datos, pero después de extraer los rezagos y las diferencias, nos quedamos con un total de 313 datos.

4.2 Variables significativas

Primeramente, buscaremos las variables significativas desde un punto de vista econométrico para después contrastarlas con la significancia, de acuerdo con el algoritmo F Test de Matlab hecho para ML, entonces procederemos a un modelo clásico de regresión lineal múltiple sin intercepto considerando a las variables explicativas con su logaritmo, a la manera de la ecuación (6):

$$\begin{aligned} \ln y_t^{MM} = & \beta_1 \ln x_t^{WTI} + \beta_2 \ln x_t^{BRENT} + \beta_3 \ln x_t^{DJI} \dots \\ & \dots + \beta_4 \ln x_t^{XAU} + \beta_5 \ln x_t^{OILC1} + \varepsilon_t \end{aligned} \quad (6)$$

Recalcando que cada una de las variables es una variable independiente en el tiempo t , según aparece en la ecuación (7):

$$\ln(x_t) = (\ln x_t^{WTI}, \ln x_t^{BRENT}, \ln x_t^{DJI}, \ln x_t^{XAU}, \ln x_t^{OILC1}) \quad (7)$$

Posterior a las variables significativas que nos arroje el modelo, usaremos esas variables para ejecutar ahora un modelo ARDL⁶³ con el fin de esclarecer en qué rezago se encuentran las variables explicativas autorregresivas con significancia en el modelo. Hay que tener en cuenta que las variables explicativas continuarán teniendo su logaritmo en adelante. Como vemos con (Nkoro & Uko, 2016), tenemos el modelo $ARDL(p, q_1, q_2, \dots, q_k)$, Donde L es el operador de rezago, $p = 0, 1, 2, \dots, m$; $q = 0, 1, 2, \dots, m$; $i = 1, 2, \dots, k$ y m es el orden de rezago máximo, además, x_{it} representan las tendencias temporales, *dummies* estacionales o las variables independientes con los rezagos fijos, según la ecuación (8):

$$\phi(L, p)y_t = \sum_{i=1}^k \beta_i (L, q_i)x_{it} + u_t \quad (8)$$

Donde:

$$\phi(L, p) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad (8.1)$$

$$\beta(L, q) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q, \quad (8.2)$$

$$\text{para } i = 1, 2, 3, \dots, k, v_t \sim iid(0; \delta^2) \quad (8.3)$$

Véase a (Nkoro & Uko, 2016) para mayor referencia.

Una vez obtenido el número de rezagos óptimo se procederá a realizar la red neuronal con ese número de rezagos, pero con todas las variables que se establecieron desde un principio.

A la manera de (Meneses-Bautista & Alvarado, 2017), en la ecuación (9) se denota a N como la función de pronóstico obtenida por la red neuronal de entrenamiento; f y g las funciones definidas anteriormente, al igual que y_t, x_{t+1}^{MM} el precio de la MM.

$$N(x_t) = g(x_t, F_t) = f(x_{t+1}) = y_t \quad (9)$$

⁶³ Autoregressive Distributed Lag.

Usando por tanto las variables explicativas de la mezcla mexicana con la estructura de (Meneses-Bautista & Alvarado, 2017) tenemos en la ecuación (10) que:

$$N(x_t) = N(x_t^{MM}, x_t^{WTI}, x_t^{BRENT}, x_t^{DJI}, x_t^{XAU}, x_t^{OILC1}) = y_t = x_{t+1}^{MM} \quad (10)$$

Recordando que as variables están transformadas en su logaritmo.

4.3 Procedimiento

Para encontrar una arquitectura ideal que modele la función N se realizan ensayos de redes con distinta cantidad de capas y neuronas. Tomando en cuenta que añadir muchas capas ocultas no beneficia al modelo y acarrea un costo computacional de acuerdo con (Gupta & Pandey, 2018), se procederá, con ayuda de *TensorFlow*, a probar 6 arquitecturas de redes neuronales, dos RNN, dos LSTM y dos BI-LSTM. Son dos de cada una porque una le corresponde a una capa oculta y otra dos capas ocultas del tipo de arquitectura en cuestión, añadido a lo anterior, todos los tipos de arquitectura terminan con una capa de salida de densidad uno, además de contener de 2 a 36 celdas (en el caso de las de doble capa oculta, el número de celdas de la primera será igual al número de celdas de la segunda). Se usará el *batch-size* por defecto de *TensorFlow*. Además, las variables independientes usadas estarán diferenciadas en 1 periodo.

Posteriormente tomaremos el RMSE más bajo de cada arquitectura de red neuronal. Tendremos un total de 6 modelos: uno para el RNN de una capa, otro para el RNN de dos capas, el siguiente para el LSTM de una capa, posteriormente el LSTM de dos capas y, por último, un BI-LSTM de una capa y el BI-LSTM de dos capas, y haremos un modelo de ML usando a la regresión lineal múltiple como referencia para establecerla como línea base⁶⁴, de esta forma obtendremos el RMSE base y las redes neuronales deben de tener un RMSE más bajo para considerarse como superiores en su desempeño en comparación con el modelo base. En este caso, el RMSE que nos compete es el de prueba y no el de entrenamiento.

Cuando se haga la regresión lineal múltiple se usarán las mismas características de entrada que las de los algoritmos de redes neuronales y se usará también 22% de los datos como datos de prueba y el resto como datos de entrenamiento.

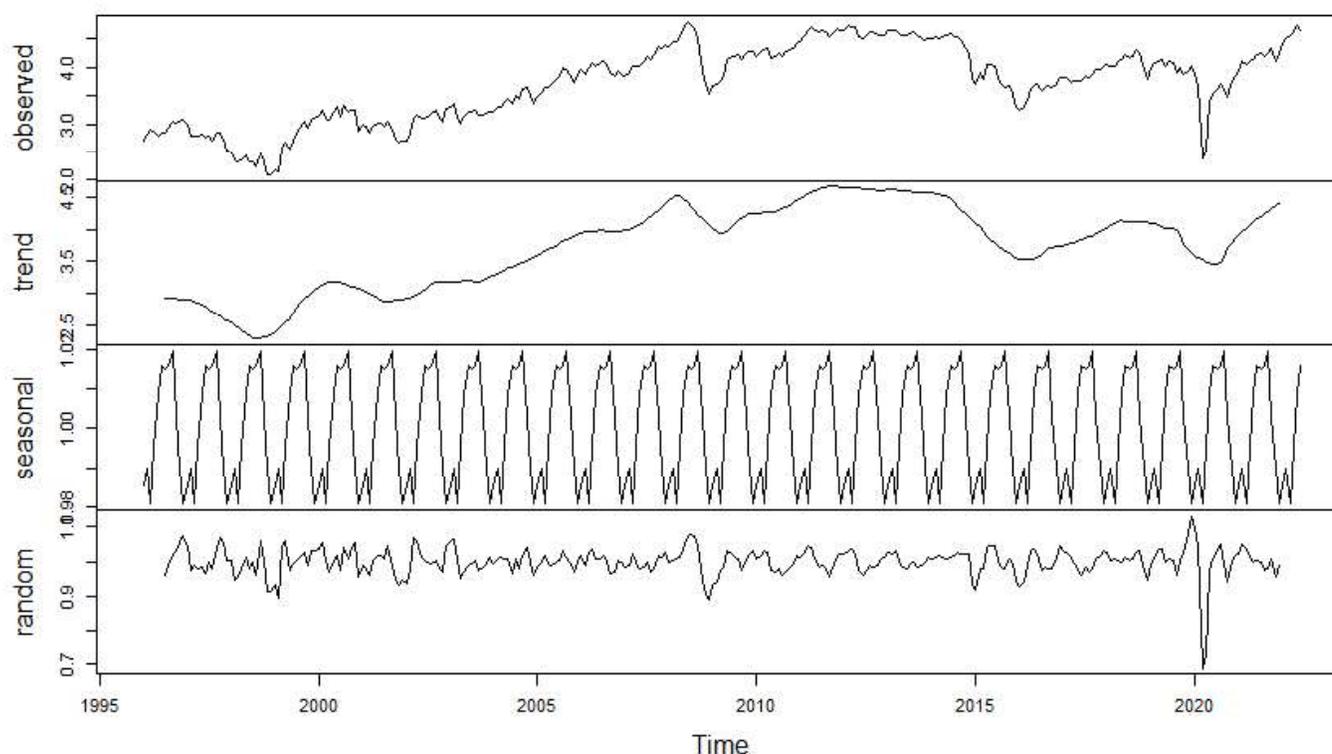
⁶⁴ Es el *base line*.

5. Análisis y resultados

5.1 Análisis

Procedimos a descomponer multiplicativamente el logaritmo natural de la serie de tiempo de la mezcla mexicana. Véase la ilustración 5.

Ilustración 5 Descomposición de los logaritmos de los precios de la Mezcla Mexicana



Fuente: Elaboración propia con datos de Banco de México.

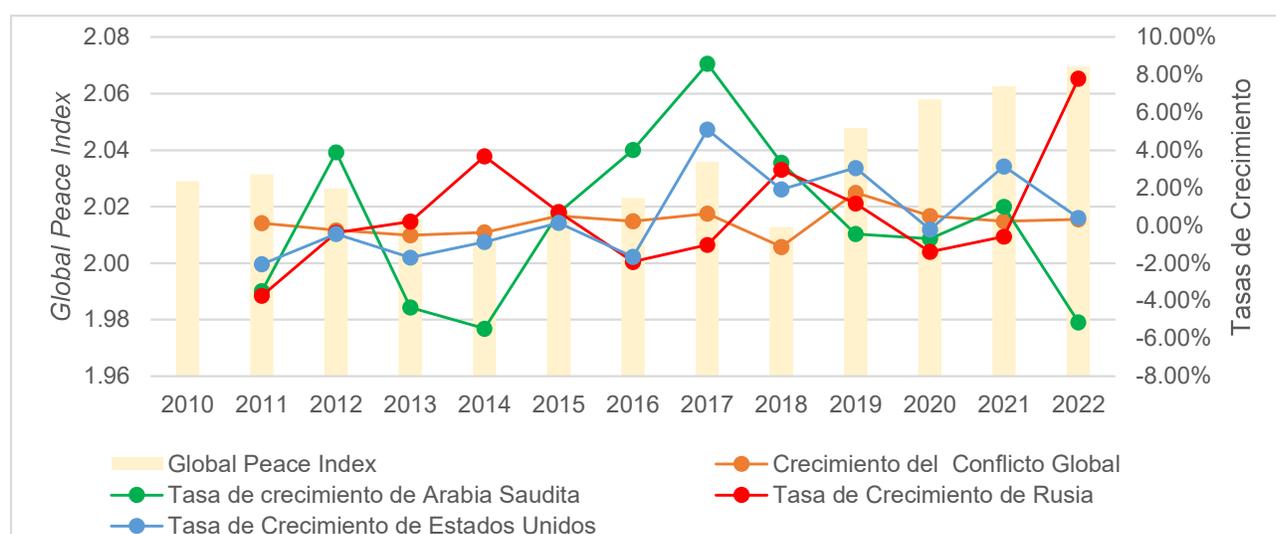
Cómo podemos ver, en los datos observados existe una caída muy fuerte en 2020 y aunque la tendencia había descendido en los meses anteriores, se recuperó posteriormente a los niveles previos a 2020, no obstante, desde antes del 2000 podemos apreciar una tendencia consistentemente al alza hasta 2007 y cae en 2008-2009 pero se mantiene estable de 2011 hasta 2014 para volver a caer en 2015 y 2016. Así mismo, a pesar de ser series mensuales, en la parte del gráfico “aleatorio”⁶⁵ vemos que en 2009, la crisis financiera afectó de igual forma al mercado energético, explicado por (Yu, Guo, & Chang, 2022) además en 2020, se puede ver que rompe con la banda de 0.9 y 1.0, por lo que podemos darnos cuenta de la alta volatilidad del petróleo de la Mezcla Mexicana, esto último lo explican (Corbet, Hou, Hu, & Oxley, 2021) que en ese periodo, en específico en abril 2020, tras un shock en la oferta de petróleo, se produjo un colapso en los precios de los derivados del WTI, llegando los precios a ser negativos. En este caso, no nos afecta en las series de tiempo porque al ser las series mensuales, los precios del OILC1 que se tomaron para esa fecha no estuvieron debajo de 0 dólares, no

⁶⁵ En la tabla está como “Random”.

obstante, indagando en los años 2015 y 2016, de acuerdo con (Ma, Xiong, & Bao, 2021), Arabia y Rusia han estado cooperando satisfactoriamente para enfrentarse a Estados Unidos en su *shale oil revolution*⁶⁶ desde 2016, además añadiendo los comentarios de (Bradshaw, Van de Graaf, & Connolly, 2019), el conflicto para el nuevo orden como potencia de producción petrolera se encuentra no solo entre Rusia y Arabia Saudita; Estados Unidos también está en este juego. Este tipo de conflictos geopolíticos podrían medirse con índices como el *Global Peace Index (GPI)*. Véase el reporte (Institute for Economics & Peace, 2022).

Aunque el reporte del GPI son datos anuales y no es posible meter este tipo de datos en las redes neuronales porque los datos en este trabajo son mensuales, si se puede apreciar que el conflicto global ha aumentado desde 2014, pues como lo indica el (Institute for Economics & Peace, 2022), el GPI utiliza 23 indicadores cualitativos y cuantitativos que se encuentran dentro de los dominios del nivel de seguridad social y protección, conflicto nacional e internacional y grado de militarización. De una escala de 1 a 5, entré mayor sea este número, mayor es el conflicto de este país. Observando la ilustración 6, Arabia Saudita sobresale por el incremento en el índice de 2015 a 2017 mientras que Rusia acentúa el nivel de conflicto en 2022.

Ilustración 6 Global Peace Index como medida de Conflicto Global (2010-2022)



Fuente: Elaboración propia con datos de <https://www.visionofhumanity.org/>.

Nota 1: Los datos de 2018 no incluyen los países Iraq, Sur de Sudán, Siria y Afganistán debido a que no se presentaron datos ese año, pero en los demás años si están incluidos.

Nota 2: Se elaboró de acuerdo con el promedio de los reportes anuales de la IEP⁶⁷ que incluye a 163 países.

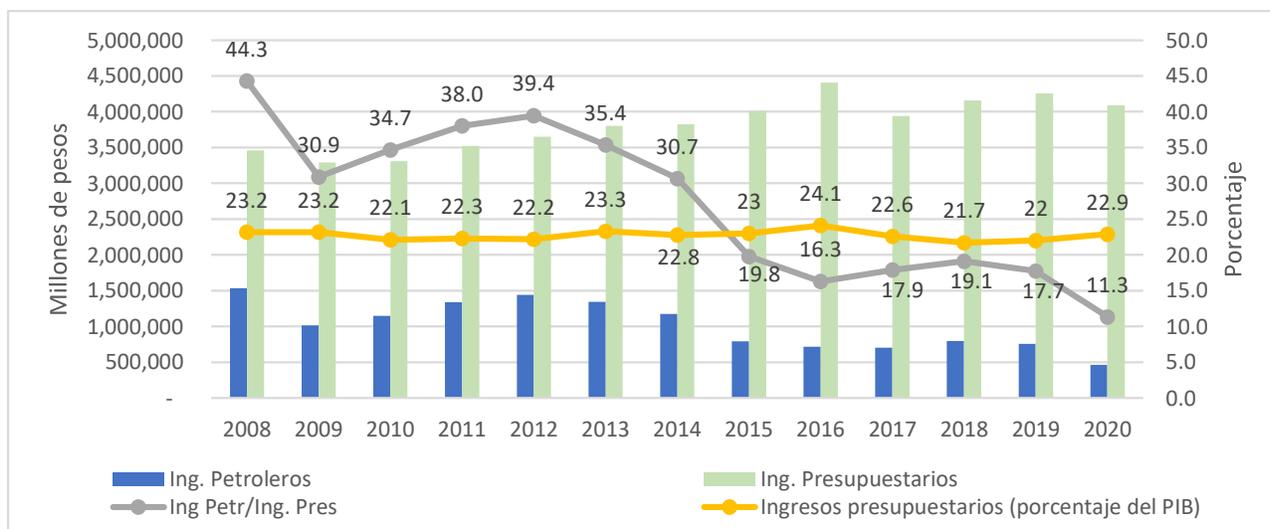
Nota 3: Palestina no se encuentra de 2015 a 2010 y el Sur de Sudán no se encuentra únicamente en 2010.M

⁶⁶ Se refiere a la combinación de fracturación hidráulica y perforación horizontal que permitió a los Estados Unidos aumentar significativamente su producción. Para más información véase la siguiente página de internet: <https://www.strausscenter.org/energy-and-security-project/the-u-s-shale-revolution/#:~:text=The%20E2%80%9CShale%20Revolution%E2%80%9D%20refers%20to,total%20U.S.%20crude%20oil%20production.>

⁶⁷ Institute for Economics & Peace.

Mientras tanto en un contexto de finanzas públicas, de acuerdo con los reportes del (Instituto Belisario Domínguez, 2017) y la (Secretaría de Hacienda y Crédito Público, 2019) los ingresos presupuestarios se han encontrado en un intervalo de 21.7% a un 24.1% con respecto al PIB⁶⁸; alcanzando el mínimo en 2018 y el máximo en 2016. Véase ilustración 7.

Ilustración 7 Ingresos petroleros con respecto a los ingresos presupuestarios del sector público (porcentaje) y los ingresos presupuestarios (porcentaje del PIB)



Fuente: Elaboración propia con información de SHCP, informes sobre la situación económica, las finanzas públicas y la deuda pública⁶⁹.

Nota 1: La proporción del PIB es base 2013 y los ingresos se encuentran en precios de 2013.

Nota 2: Los factores de INPC que se usaron fueron tomados de los índices de diciembre de cada año.

Como se aprecia en la ilustración 7, los ingresos petroleros con respecto a los ingresos presupuestarios del sector público han disminuido desde 2012, alcanzando su mínimo en 2020 con apenas 11.3%, de acuerdo con los reportes de (Instituto Belisario Domínguez, 2017) y de (Secretaría de Hacienda y Crédito Público, 2021). Por otro lado, a pesar de que la caída de precios de la mezcla mexicana alcanzó un precio de 37⁷⁰(dpb⁷¹) en 2016⁷², los ingresos presupuestarios tuvieron un máximo histórico tomando en cuenta el periodo de 2008 a 2020⁷³. Además de las referencias anteriores, véase los reportes de (Secretaría de Hacienda y Crédito Público, 2020), (Secretaría de Hacienda y Crédito Público, 2019) y (Secretaría de Hacienda y Crédito Público, 2018)

En el mismo periodo de 2008 a 2020, los años en que se ejercieron las coberturas petroleras fueron en los años de 2009, 2015, 2016 y 2020, de los cuales de acuerdo con

⁶⁸ Producto Interno Bruto.

⁶⁹ Para consultar la información puede dirigirse a la siguiente página de internet:

https://www.finanzaspublicas.hacienda.gob.mx/es/Finanzas_Publicas/Informes_al_Congreso_de_la_Union

⁷⁰ Promedio mensual y elaborado con datos de Banco de México.

⁷¹ Dólares por barril.

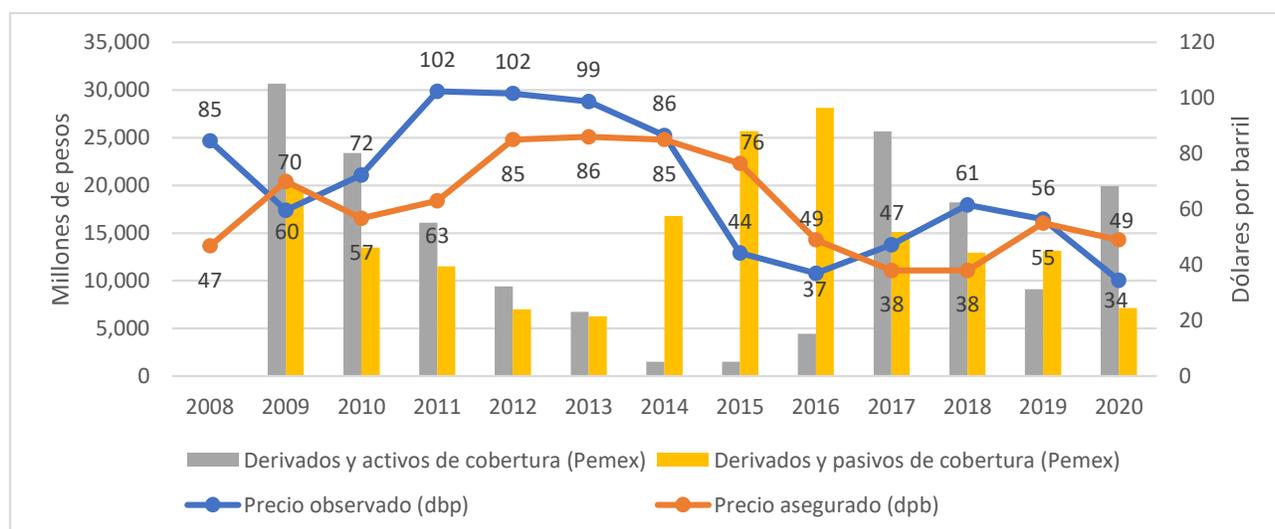
⁷² El segundo nivel más bajo, pues el nivel más bajo se encuentra en 2020 con 34(dpb).

⁷³ Véase ilustración 8.

el (Centro de Estudios de las Finanzas Públicas, 2020) se aseguraron a un precio de 70, 76.4, 49 y 49 dólares por barril respectivamente cuando en el mercado el precio promedio mensual observado⁷⁴ se encontraba en 60, 44, 37 y 34 dólares por barril respectivamente. (Véase ilustración 8).

De igual forma, como se ve en la ilustración 8, los activos de cobertura de Pemex tomados del balance general⁷⁵ fueron disminuyendo desde 2009 hasta 2015 mientras que los pasivos de cobertura tuvieron un alza importante en los periodos de 2014 a 2016, periodos en los cuales se tuvo la caída más pronunciada de la mezcla mexicana, de acuerdo con el (Instituto Belisario Domínguez, 2017). Pasó de estar en 2014 con un promedio mensual de 86 dólares por barril⁷⁶ a un promedio mensual de 37 dólares por barril⁷⁷ en 2016, lo que representa una caída en dicho periodo de 56.98%, no obstante, como se ve en la ilustración 7, los ingresos presupuestarios no disminuyeron y además, la proporción de los ingresos petroleros con respecto a los ingresos presupuestarios disminuyó.

Ilustración 8 Activos y pasivos de cobertura de Pemex, así como el precio promedio mensual observado y precio asegurado (dólares por barril)



Fuente: Elaboración propia con datos del CEFP⁷⁸, Bloomberg L.P. y Banco de México.

Nota 1: Los activos y pasivos de coberturas de Pemex se encuentran en precios de 2013.

Nota 2: Los factores de INPC usados para llevarlos a precios de 2013 fueron tomados de los índices de diciembre de cada año.

Nota 3: Los activos y pasivos de cobertura fueron extraídos del balance general estandarizado.

Aunado a los párrafos anteriores, los ingresos y gastos por las coberturas petroleras se pueden ver con un mejor contexto, ya que, de acuerdo con el (Centro de Estudios de las

⁷⁴ Elaborado con datos de Banco de México.

⁷⁵ Información extraída de Bloomberg L.C.

⁷⁶ Información extraída de Banco de México.

⁷⁷ Información extraída de Banco de México.

⁷⁸ Centro de Estudios de las Finanzas Públicas. Los datos se pueden consultar en la siguiente liga:

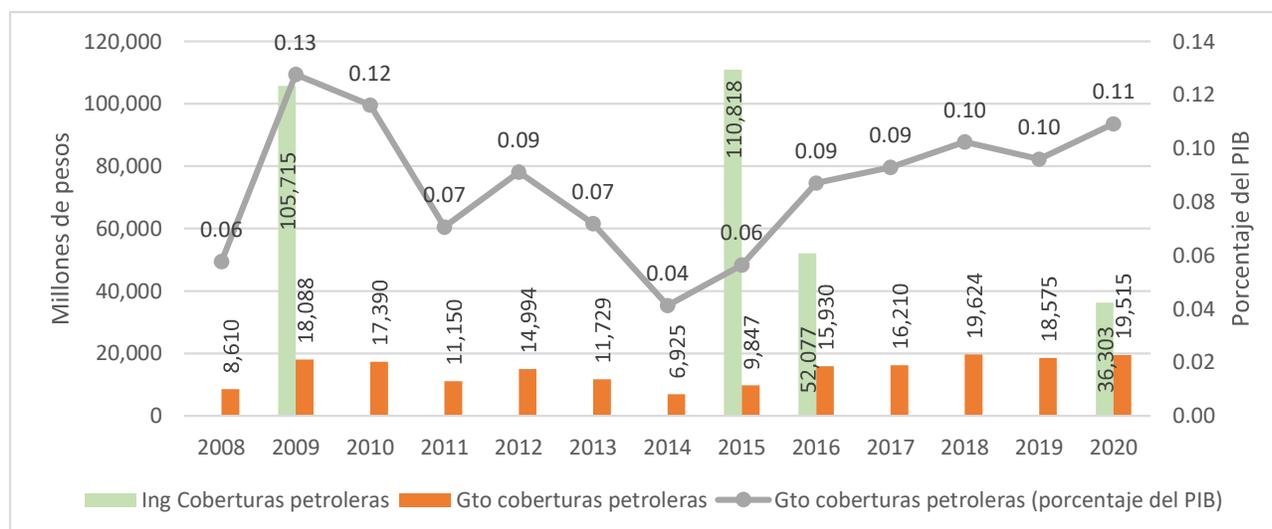
<https://www.cefp.gob.mx/publicaciones/nota/2021/notacefp0132021.pdf>

Finanzas Públicas, 2020) de 2008 a 2022, el gasto representó 207,642 millones de pesos con un ingreso 312,944 millones de pesos, los cuales, en precios de 2013 representa 188,586 y 304,913 millones de pesos respectivamente⁷⁹. Esto quiere decir que hubo una variación porcentual de 33.6% en términos corrientes, aunque a precios de 2013 ascienda a 37.1%.

Por otro lado, como se puede ver en la ilustración 9, el gasto de las coberturas petroleras tiene un intervalo mínimo con respecto al PIB⁸⁰ de 0.04%⁸¹ alcanzado en 2014 mientras que el intervalo máximo lo situamos en 2009 con un 0.13% con respecto al PIB⁸².

Esto quiere decir que podría evitarse una parte del gasto de coberturas bajo el supuesto de que los pronósticos de redes neuronales fueran lo más parecido al precio observado y se evitara gastar en coberturas petroleras⁸³ en los años en los que el pronóstico fuera más alto que el precio al que se pudiera pactar en el mercado. Bajo el supuesto anterior, el impacto que hubiera tenido de 2008 a 2020 sería de un ahorro máximo de 207,642 millones de pesos, puesto que, ese fue el gasto total por coberturas cambiarias durante esos años según el (Centro de Estudios de las Finanzas Públicas, 2020).

Ilustración 9 Ingreso y gasto de las coberturas petroleras (millones de pesos de 2013)



Fuente: Elaboración propia con datos del CEFP e información de SHCP, informes sobre la situación económica, las finanzas públicas y la deuda pública.

Nota 1: La proporción del PIB es base 2013.

Nota 2: Los ingresos y gastos de las coberturas cambiarias se encuentran a precios de 2013 y el INPC utilizado para calcular el factor de actualización es al índice de diciembre de cada año.

⁷⁹ Los ingresos y gastos por coberturas petroleras a precios de 2013 pueden verse en la ilustración 9.

⁸⁰ Base 2013.

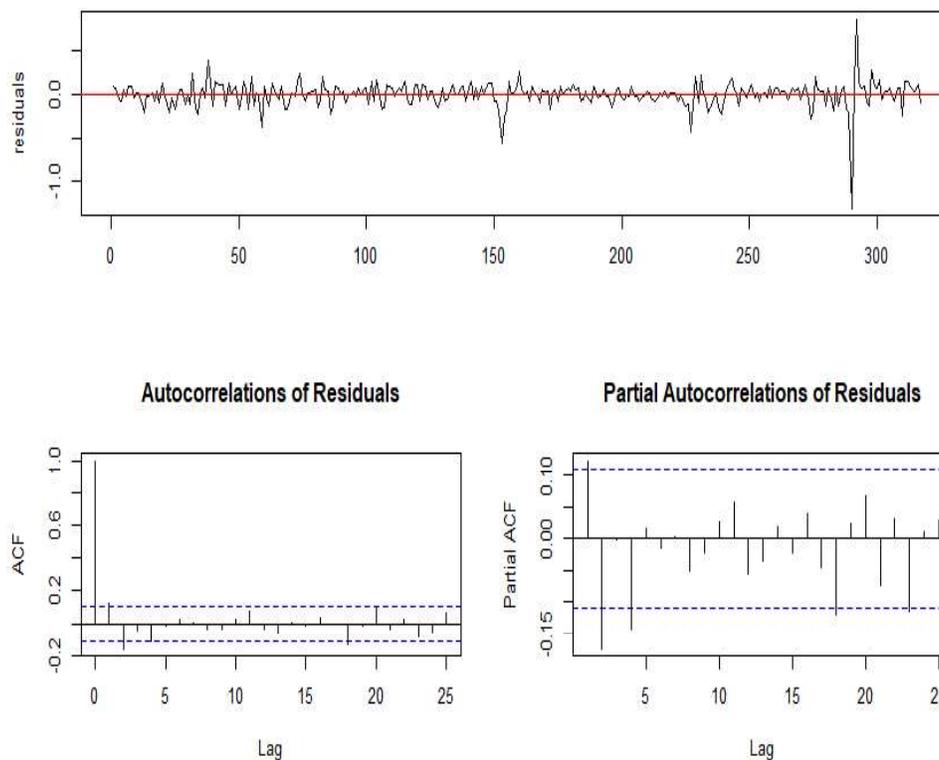
⁸¹ Recordar: Los montos se encuentran a pesos de 2013.

⁸² Base 2013.

⁸³ O al menos disminuirlas.

Volviendo a los datos estadísticos, en la ilustración 10 aparecen los residuales y su autocorrelación. Destaca el rezago 2 y 4 de la correlación parcial.

Ilustración 10 Autocorrelación y Autocorrelación parcial de los residuales de los logaritmos de los precios de la Mezcla mexicana



Fuente: Elaboración propia con datos DE Banco de México.

Aunque en la prueba de Dickey-Fuller Aumentada de la Tabla 1 podemos prever que algún modelo con 4 rezagos podría ser significativo, usando el logaritmo y la primera diferencia de la serie, vemos que la serie **no** es estacionaria pues existe un *p-value* a 0.01, lo cual indica que no rechazamos la hipótesis nula de no estacionariedad.

Tabla 1 Prueba Dickey-Fuller de los logaritmos de los precios de la Mezcla Mexicana

Augmented Dickey-Fuller Test
data: tsstationary
Dickey-Fuller = -8.9458, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Fuente: Elaboración propia con datos de Banco de México.

Después de esto, procedemos a hacer el modelo con 4 rezagos y nos da como resultado un SARIMA (4,1,4) (3,0,3) muy adecuado, ya que nos permite enfocarnos en las variables rezagadas significativas que se encuentran en la Tabla 2.

Tabla 2. Z test de Coeficientes de la serie logarítmica de precios de la Mezcla Mexicana

Variable	Estimate	Std. Error	z value	Pr(> z)
ar1	0.62337183	0.09312634	6.6938	0.0000000002174 ***
ar2	-0.81458412	0.03657845	-22.2695	<0.0000000000000022***
ar3	0.28216957	0.06522695	4.326	0.00001519***
ar4	0.15767236	0.01466731	10.7499	<0.0000000000000022***
ma1	-0.4862665	0.23704972	-2.0513	0.0402351*
ma2	0.61282772	0.34242383	1.7897	0.073506
ma3	-0.08303857	0.13426603	-0.6185	0.5362702
ma4	-0.44479221	0.33376855	-1.3326	0.1826511
sar1	-0.2431545	0.07096796	-3.4263	0.000612***
sar2	0.25283246	0.07430642	3.4026	0.0006676***
sar3	0.97774798	0.00072775	1343.5178	<0.0000000000000022***
sma1	0.27902076	0.15575749	1.7914	0.0732324
sma2	-0.28721002	0.2158379	-1.3307	0.1832961
sma3	-0.95083523	0.07575906	-12.5508	<0.0000000000000022***

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fuente: Elaboración propia con datos de Banco de México.

Como podemos ver en la Tabla 2, existe significancia en los rezagos 4, 3, 2 y 1 de la mezcla mexicana. Por tanto, podemos aceptar, desde el punto de vista estadístico, que es válido el uso de 4 rezagos para incorporar en el algoritmo de la red neuronal. Ahora, debemos saber si existe significancia en las demás variables explicativas.

Alternativamente, diseñamos una regresión lineal múltiple usando las 5 variables explicativas y la variable de la Mezcla Mexicana, que es la variable dependiente. Para ello, empleamos los logaritmos naturales de las diferencias en los precios. Los resultados de la regresión se muestran en la Tabla 3.

Tabla 3. Regresión Múltiple

Variable	Estimate	Std. Error	t value	Pr(> t)
WTI_In_diff	-0.10539	0.1994	-0.529	0.5975
BRENT_In_diff	0.67676	0.03513	19.265	<0.0000000000000022***
DJI_In_diff	0.08849	0.04705	1.881	0.06093
XAU_In_diff	0.06966	0.04354	1.6	0.11067
OILC1_In_diff	0.54758	0.19699	2.78	0.00577**

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

En este caso, las variables que más significancia tienen para la Mezcla Mexicana son el BRENT y el Contrato 1 de futuros de petróleo *Light-Sweet*, por lo que, en el modelo de ARDL, ensayamos hasta 4 rezagos en el Brent y el OILC1, sin embargo, cuando se

realizó el análisis de coeficientes, solo el BRENT hasta con 4 rezagos tuvo significancia, aunque sí pudimos ver significancia en todas las variables sin rezago. Véase Tabla 4.

Tabla 4. Modelo ARDL

Variable	Estimate	Std. Error	t value	Pr(> t)
L(MM_In, 1)	0.650932	0.055814	11.663	< 2e-16 ***
L(MM_In, 2)	0.106007	0.055366	1.915	0.056474 .
BRENT_In	0.934032	0.03172	29.446	< 2e-16 ***
L(BRENT_In, 1)	-0.60038	0.05949	-10.09	< 2e-16 ***
L(BRENT_In, 2)	-0.20156	0.05731	-3.517	0.000503 ***
L(BRENT_In, 3)	0.041154	0.02375	1.733	0.084151 .
L(BRENT_In, 4)	-0.03855	0.016364	-2.356	0.019127 *
DJI_In	-0.02612	0.004549	-5.742	2.27e-08 ***
XAU_In	0.033906	0.007488	4.528	8.56e-06 ***
OILC1_In	0.104989	0.033358	3.147	0.001811 **

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

Por tanto, teniendo significancia hasta el cuarto rezago de la Mezcla Mexicana y del BRENT, será válido usar hasta 4 rezagos en las variables de entrada cuando se realicen los algoritmos de ML, no obstante, hay que recordar que esto es solo un modelo econométrico. En caso de que las variables no fueran significativas desde el punto de vista de ML, es ahí donde se tendría que replantear el número de rezagos de manera independiente. De esta manera, tendremos en cuenta la selección del BRENT, DJI, XAU y OILC1 en su logaritmo natural y sus diferencias hasta su cuarto rezago al igual que la misma Mezcla Mexicana en el rezago 4, de igual forma consideramos WTI porque, aunque no se consideró en el modelo, nos basta saber que existen hasta 4 rezagos significativos y que la mayoría de las variables son significativas. Estas variables serán usadas en la capa de entrada para pronosticar la Mezcla Mexicana de la siguiente observación, sin embargo, usaremos el *Feature Ranking Algorithm*⁸⁴ de Matlab para saber si, desde el punto de vista de ML, las variables que elegimos efectivamente tienen importancia para predecir.

⁸⁴ Algoritmo que se va a usar en Matlab para ver las variables significativas.

5.2 Resultados

5.2.1 Modelo Base

El *modelo base* de ML en el que se emplea la regresión múltiple nos arrojó los resultados de la Tabla 5.

En el Caso de Matlab se eligieron el 22% de las observaciones, lo cual, en Matlab representaron 68 observaciones para pruebas y el resto se usó para entrenamiento, de igual forma se usa el logaritmo natural y sus diferencias para tener la variable respuesta de la Mezcla Mexicana, además de 4 rezagos de las respectivas variables independientes con un tipo de validación: *5-fold cross-validation*. Nos deja con un RMSE a superar de 0.0398 en validación y un RMSE de pruebas en 0.0411⁸⁵ Por otro lado, aunque no estamos evaluando en este momento el R cuadrado, sí podemos percatarnos de que es muy alto, en un 0.8819 para las pruebas.

Tabla 5. Modelo de Machine Learning usando una Regresión Lineal Múltiple como modelo base. 4 rezagos

Session: RegressionLearnerSession
 Training Data: resumenpetroleo4rezagoslnhydifff Observations: 245
 Predictors: 29 Predictor Names: WTI, BRENT, DJI, XAU, OilC1, MM1R, WTI1R, BRENT1R, DJI1R, XAU1R, OilC11R, MM2R, WTI2R, BRENT2R, DJI2R, XAU2R, OilC12R, MM3R, WTI3R, BRENT3R, DJI3R, XAU3R, OilC13R, MM4R, WTI4R, BRENT4R, DJI4R, XAU4R, OilC14R
 Response Name: MM
 Validation: 5-fold cross-validation
 Test Data: ResumenpetroleoIndiff Observations: 68

Model Type	Status	RMSE (Validation)
Linear Regression	Tested	0.0398
RSquared (Validation)	RSquared (Test)	RMSE (Test)
0.9287	0.8819	0.0411

Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

⁸⁵ El RMSE de pruebas es la que nos compete en el presente trabajo.

5.2.2 Variables significativas usando el algoritmo F Test para modelos de *Machine Learning*

Para comprender el nombre de las variables, el código que se encuentra después del nombre: 1R, 2R, 3R y 4R representan los rezagos 1, 2, 3 y 4 respectivamente.

La prueba F examina la importancia de cada predictor y luego clasifica las características usando el *p-value*(p). Su hipótesis alternativa es que las medias de población no son todas iguales y el *score*⁸⁶ de la prueba F corresponde al $-\log(p)$. i.e. entre más bajo es el *p-value*, la prueba indica que la variable es más importante. Si el *p-value* es menor que un ϵ (0)⁸⁷ entonces el resultado es Inf⁸⁸. Por consiguiente, sabiendo que el *p-value* solo puede tomar valores entre 0 y 1, cuando el *p-value* es 1, entonces la prueba F es cero porque el logaritmo de 1 es 0. Por tanto, un *p-value* cercano a 1 sugiere una prueba F de 0, lo que significa que no hay diferencia entre los grupos. (Matlab, 2022).

Cabe señalar que, entre los mejores resultados de acuerdo con el nivel de importancia de la clasificación se encuentran al BRENT en primer lugar, seguido por el OILC1 y el WTI con resultados mayor a 80 en la prueba F. Véase tabla 6.

Si bien, todas las variables tienen algo que aportar porque ninguna tiene un F test de cero, al tomar en cuenta un *p-value* de 0.05, la prueba F nos indica un valor de 1.3010, por tanto, solo las primeras 15 variables tendrían significancia estadística⁸⁹. Dentro de esas 15 variables podemos encontrar significancia de hasta el cuarto rezago en algunos casos, por lo tanto, se procederá a usar las variables independientes establecidas previamente hasta su cuarto rezago.

⁸⁶ Puntuación.

⁸⁷ $2.2204e-16$.

⁸⁸ Infinito positivo.

⁸⁹ Teniendo en cuenta que la prueba F corresponde al $-\log(p)$.

Tabla 6 F Test. Resultados del algoritmo

Place	Features	F Test	Place	Features	F test
1	BRENT	87.9994	16	MM3R	0.9225
2	OILC1	86.6040	17	DJI2R	0.6633
3	WTI	83.2367	18	MM2R	0.5798
4	DJI	6.6706	19	XAU4R	0.5212
5	XAU	5.1808	20	MM4R	0.4929
6	MM1R	4.5585	21	DJI3R	0.4014
7	WTI1R	3.1106	22	WTI3R	0.3094
8	OiIC11R	2.8845	23	OiIC13R	0.2601
9	DJI4R	2.3625	24	XAU2R	0.2518
10	OiIC14R	2.3470	25	XAU1R	0.1602
11	XAU3R	2.1393	26	BRENT3R	0.11
12	BRENT4R	2.0978	27	WTI2R	0.0939
13	DJI1R	1.9705	28	BRENT2R	0.0721
14	WTI4R	1.3213	29	OiIC12R	0.0334
15	BRENT1R	1.3103			

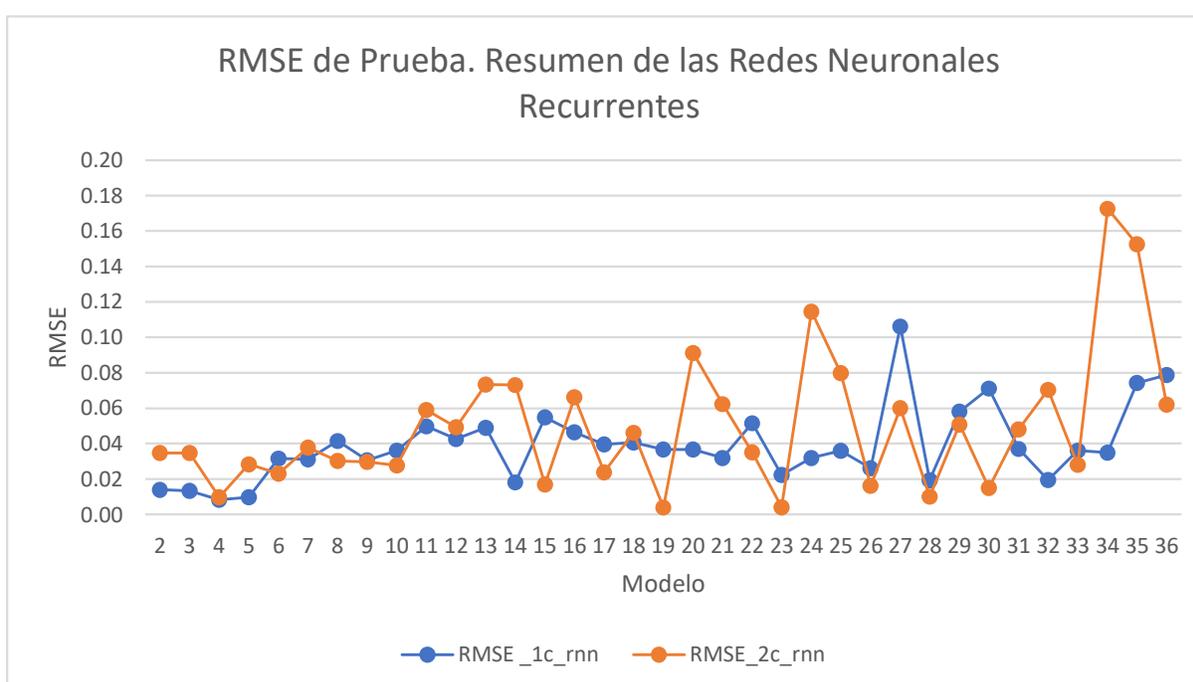
Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

5.2.3 Redes Neuronales

5.2.3.1 Redes Neuronales Recurrentes

Teniendo en cuenta que se debe obtener un modelo de Redes Neuronales mejor al RMSE de testeo de 0.0411 del modelo base de regresión, se prosiguió a realizar el entrenamiento de las redes y, en consecuencia, se corrieron 210 modelos. Así se obtuvo el mejor de cada configuración de los 6 diferentes algoritmos utilizados, es decir, en total fueron 35 por algoritmo, pero antes de analizar los resultados debemos percatarnos que en la ilustración 11 el acrónimo 1c y 2c se refieren a 1 capa y 2 capas.

Ilustración 11 Resumen del desempeño de Redes Neuronales recurrentes



Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

En la ilustración 11 podemos ver que el RMSE de los modelos es más bajo conforme menos neuronas tengan, el mejor modelo RNN de dos capas es el 19⁹⁰ con un RMSE de 0.0040 y el mejor del modelo RNN de una sola capa oculta fue el modelo 4⁹¹ con un error de 0.0083; por otro lado, el algoritmo se vuelve inestable a partir del modelo RNN 13⁹² de dos capas ocultas, no obstante, si consideramos que el RMSE del modelo base fue 0.0411, ambos se encuentran por debajo de ese nivel, además varios modelos de redes neuronales son mejores que ese nivel si vemos la ilustración 11.

⁹⁰ Es el modelo RMSE19b.

⁹¹ RMSE4: Significan 4 celdas en la capa oculta. El modelo es un RNN de una sola capa.

⁹² RMSE13: 13 celdas en la capa oculta.

Cómo vemos en la ilustración 11, en la red neuronal “2c_rnn”, existe una estabilidad antes del modelo RMSE10b⁹³, no obstante, no es sino hasta el modelo RMSE19b⁹⁴ donde se alcanza el mejor modelo con un RMSE de 0.0040.

Comparando el RNN de 1 capa oculta y el RNN de 2 capas ocultas, encontramos que el RNN de 2 capas suele tener más saltos en cuanto al desempeño según cambiemos el número de celdas en cada capa oculta, no obstante, a pesar de ello, el algoritmo de 2 capas oculta tiene el mejor desempeño de los 2.

⁹³ RMSE10b: 10 celdas en cada capa oculta.

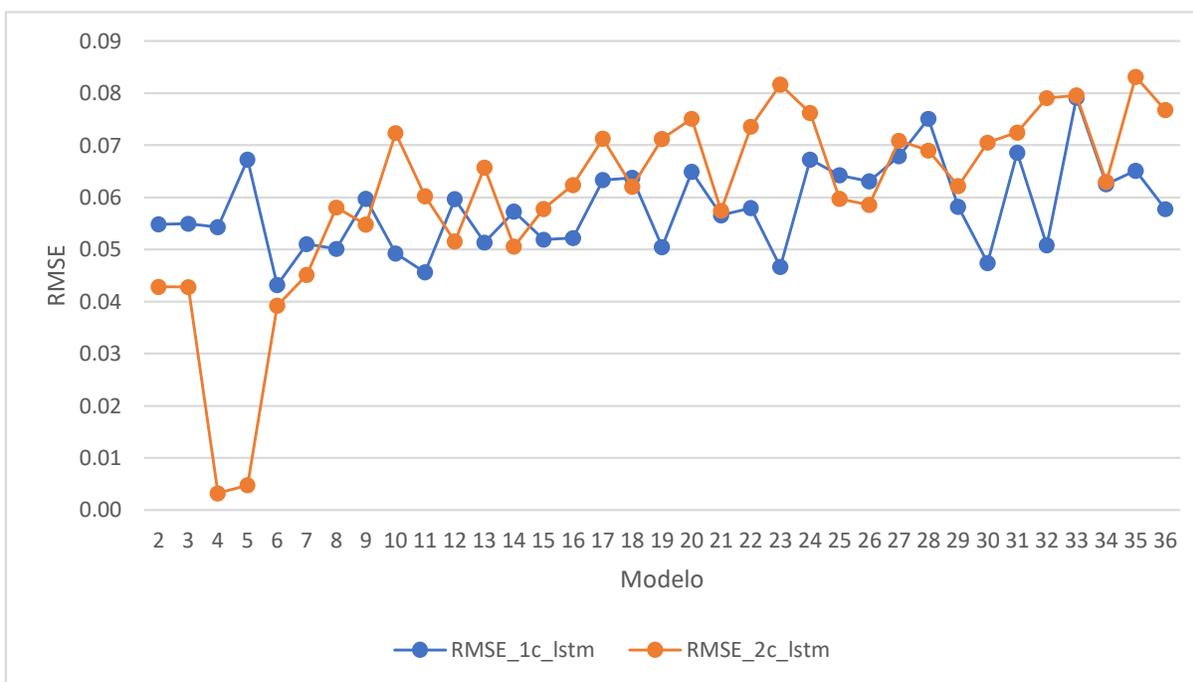
⁹⁴ RMSE19b: 19 celdas en cada capa oculta.

5.2.3.2 Redes Neuronales LSTM

De la misma forma que las redes neuronales recurrentes, se corrieron 35 configuraciones distintas dentro del algoritmo LSTM para una capa oculta y otras 35 configuraciones con dos capas ocultas, cambiando el número de celdas en cada capa oculta, desde 2 hasta 36, posteriormente comparamos el algoritmo de LSTM de una sola capa con respecto al de dos capas.

En la ilustración 12, tenemos un comportamiento inverso al anterior cuando comparamos el algoritmo LSTM de 1 capa oculta con respecto a la de 2 capas ocultas. De ambos, los mejores modelos se encuentran en el algoritmo de 2 capas, en específico en el RMSE4d y RMSE5d⁹⁵, con 0.0032 y 0.0047 respectivamente, mientras que el LSTM de 1 capa oculta no goza de ningún modelo mejor que el modelo base, pues su mejor modelo tiene un RMSE de 0.0432⁹⁶

Ilustración 12 Resumen del desempeño de las redes neuronales LSTM



Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

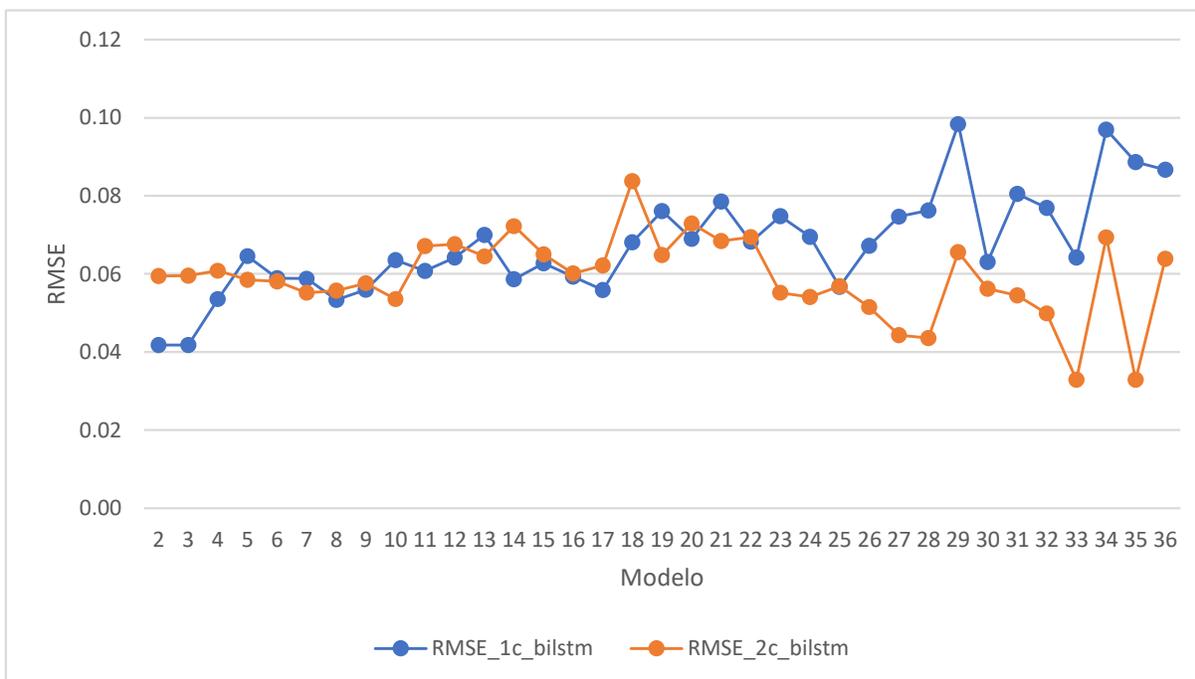
⁹⁵ 4 y 5 celdas en cada capa oculta para ambos modelos respectivamente.

⁹⁶ El modelo en cuestión es el RMSE6c, lo cual significa que tiene 6 celdas en la capa oculta.

5.2.3.3 Redes Neuronales BI-LSTM

Por último, en la ilustración 13 de modelos bidireccionales, no nos encontramos con ningún modelo que supere el *modelo base* a excepción de los últimos modelos generados por el algoritmo de 2 capas BI-LSTM, el cual, tiene por mejor el modelo RMSE33f⁹⁷ con un desempeño de 0.0329, mientras que el algoritmo de una sola capa nos arroja apenas en su mejor desempeño, un RMSE de 0.0418 (RMSE3e).

Ilustración 13 Resumen de desempeño de las redes neuronales LSTM Bidireccionales



Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

⁹⁷ MRSE33f: 33 celdas en ambas capas ocultas.

5.2.4 Resumen

Los resultados de la Tabla 7 nos indican que, los mejores modelos en su conjunto son los recurrentes de 1 y 2 capas ocultas, pero el LSTM de 2 capas ocultas tiene el modelo con mejor desempeño de todas las arquitecturas planteadas. Por otro lado, el BI-LSTM de 2 capas también pasa la prueba de estar por debajo del modelo base, pero no tiene un mejor desempeño que los anteriores mencionados.

Tabla 7 Resumen de desempeño de los 6 mejores modelos

Mejores Modelos		
Algoritmo	Capas ocultas	
	1	2
RNN	RMSE4 con 0.0083	RMSE19b con 0.0040
LSTM	RMSE6c con 0.0432	RMSE4d con 0.0032
BI-LSTM	RMSE3e con 0.0418	RMSE33f con 0.0329

Fuente: Elaboración propia con datos de Banco de México, la EIA e investing.com.

Por último, la estructura del código, así como los resultados expuestos en el presente trabajo siguen la lógica de la Tabla 8. Cabe señalar que la numeración 2 al 36 se refiere al número de celdas en cada capa oculta y la serie numérica sin letra⁹⁸ es la red neuronal de una capa oculta, así mismo, la serie la b, c, d, e y f se refieren a los modelos RNN de dos capas, LSTM de una capa, LSTM de dos capas, BI-LSTM de una capa y BI-LSTM de dos capas respectivamente.

Tabla 8 Estructura de cada red neuronal

Estructura de los modelos		
Algoritmo	Capas ocultas	
	1	2
RNN	Modelo 2-36	Modelo 2b-36b
LSTM	Modelo 2c-36c	Modelo 2d-36d
BI-LSTM	Modelo 2e-36e	Modelo 2f-36f

Fuente: Elaboración propia.

Para visualizar el código usado en el apartado de análisis y resultados, diríjase al siguiente repositorio:

<https://github.com/a01595093/tesina.git>

⁹⁸ i.e. La serie de los modelos del 2 al 36 que no poseen ninguna letra del lado derecho del número.

6. Conclusiones y recomendaciones

6.1 Conclusiones

Se comprobaron que, para los casos analizados, las Redes Neuronales Recurrentes tienen un mejor desempeño para los pronósticos que el algoritmo de la Red Neuronal LSTM y BI-LSTM⁹⁹ en su conjunto, además, de los mejores modelos seleccionados de cada algoritmo de redes neuronales, las de dos capas ocultas fueron mejores que las de una sola capa oculta, pues la red LSTM y BI-LSTM de una sola capa oculta no fueron mejores que el *base-line*.

Los modelos generados por el algoritmo LSTM fueron mejores entre menos celdas se tenían en cada capa oculta mientras que, por otro lado, en el BI-LSTM, los mejores modelos se obtuvieron con una mayor cantidad de celdas en cada capa.

Entre el algoritmo de la red neuronal recurrente de 1 y 2 capas ocultas, los resultados de una sola capa suelen ser más estables cuando se aumenta el número de celdas en cada capa oculta, no obstante, aunque el de dos capas presenta mayor inestabilidad a partir de 14 celdas, también es cierto que tuvieron un mejor desempeño que cualquier configuración del algoritmo para los modelos en los que existe una sola capa.

A pesar de que la regresión lineal múltiple nos estableció que el BRENT y el OILC1 eran las variables significativas para explicar las variaciones en la Mezcla Mexicana, cuando usamos el algoritmo F test, que está más enfocado a encontrar las variables significativas en ML¹⁰⁰, e hicimos el *base-line* de la regresión lineal múltiple para obtener el RMSE que la red neuronal debía vencer, pudimos observar que, en realidad, todas las variables que establecimos desde el principio aportan al pronóstico.

Desde el punto de vista del modelo ARDL, el rezago 4 del WTI es el único que a ese nivel de rezago es significativo, no obstante, en el F Test, se demostró que todas las variables hasta su cuarto rezago mostraban una aportación al modelo de ML en la que se usó a la regresión lineal múltiple con 4 rezagos como referencia. Ésta fue hecha el modelo base. Con base en lo anterior dicho, se probaron a las redes neuronales con 4 rezagos en las variables de entrada con sus respectivas variables independientes y se logró obtener 4 modelos mejores que el modelo base de un total de 6 resultados posibles, aunque los desempeños del algoritmo dependen de la calidad de los datos.

Una mejor estimación de los precios del petróleo podría ahorrar recursos debido a que disminuiría o evitaría gastos de las coberturas; por ejemplo, si se hubiera usado esta herramienta en el periodo de 2008 a 2020, podría haberse evitado un gasto acumulado en coberturas petroleras por hasta 207,642 millones de pesos¹⁰¹.

⁹⁹ Anteriormente referenciado como BI-LSTM.

¹⁰⁰ Anteriormente referenciado como ML.

¹⁰¹ La estimación de los gastos en coberturas petroleras se encuentra directamente en la publicación del (Centro de Estudios de las Finanzas Públicas, 2020).

La hipótesis se cumple en cuatro de los seis algoritmos de redes neuronales debido a que solo en estas cuatro, su RMSE fue menor que el modelo base planteado¹⁰². Estas 4 son: RNN de una y dos capas ocultas más LSTM y BILSTM de dos capas ocultas.

¹⁰² El modelo base fue la regresión lineal múltiple.

6.2 Recomendaciones

El *batch-size* que se usó fue el que aparece por defecto en *TensorFlow* de Python, por tanto, para un análisis más profundo, debería probarse con diferentes *batch-size*, aunque el costo computacional es mucho mayor, porque en este caso se generaron 210 modelos, pero si probamos con 5 diferentes *batch-size*, por poner un ejemplo, se tendrían que correr más de 1,000 modelos.

La selección de 4 rezagos fue determinada desde un punto de vista estadístico, no obstante, usar más rezagos podría mejorar el desempeño en las diferentes configuraciones de las redes neuronales, además, también puede ser tema de análisis la posibilidad de estandarizar los logaritmos que se generaron para observar los RMSE de esos modelos y ver si la estandarización mejora el desempeño.

Si se aumenta, en su caso, el número de rezagos. Se recomendaría probar redes con una profundidad mayor a 2 capas ocultas e inclusive hacer alguna combinación entre algún tipo de capa oculta y otro tipo de capa oculta distinta.

El próximo paso es hacer un pronóstico no solo del tipo *one-step* como el que se genera en este trabajo, sino que también es posible determinarse un tipo de pronóstico “multi-step¹⁰³” aunque para ello es necesario pronosticar las variables que se determinaron como independientes para que, posteriormente, esos datos pronosticados se puedan meter como entradas en la primera capa de las redes neuronales.

Por la facilidad de generalización de los algoritmos de redes neuronales, se recomienda su aplicación con otras variables de pronóstico de diferentes mercados, ya que, debido al uso de TensorFlow, es más fácil el cambio de los hiperparámetro, así como del número de variables de entrada o de rezagos que se puedan introducir al entrenamiento, además, usar a las medias móviles para suavizar las series de tiempo podría promover a modelos con mejores desempeños.

Se recomendaría usar variables que midan directamente la volatilidad como el índice EVL promovido por el *CME Group*, además, para tener presente variables geopolíticas, podrían tomarse en cuenta datos proporcionados por el *World Energy Outlook* o el *Institute for Economics & Peace*, pero para ello sería necesario tener otra escala de periodicidad que no sea mensual y que pueda ajustarse a los datos proporcionados por estas instituciones.

En caso de que las técnicas de pronóstico como las redes neuronales lleven a resultados más eficientes, los costos que conlleva la contratación de opciones financieras podrían disminuir significativamente sino es que anular en algún momento su contratación en periodos en los que el pronóstico nos muestre escenarios alcistas en los mercados financieros, dicho de otro modo, los costos de las coberturas petroleras disminuirían debido a la menor contratación de opciones financieras, gracias a un mejor pronóstico en los precios de petróleo.

¹⁰³ Que se puede determinar un pronóstico de varias observaciones adelante y no solamente de una observación.

Bibliografía

- Abadi, M., Isard, M., & Murray, D. G. (2017). A Computational Model for TensorFlow: An Introduction. *Association for Computing Machinery*, 1-7. Obtenido de <https://doi.org/10.1145/3088525.3088527>
- Banco Mundial. (29 de 10 de 2022). *Crecimiento del PIB (% Anual) - México*. Obtenido de Datos sobre las cuentas nacionales del Banco Mundial y archivos de datos sobre cuentas nacionales de la OCDE.: <https://datos.bancomundial.org/indicador/NY.GDP.MKTP.KD.ZG?locations=MX>
- Baron, D. (2019). Machine Learning in Astronomy: A practical overview. *arXiv Cornell University*, 37. Obtenido de <https://doi.org/10.48550/arXiv.1904.07248>
- BBC News Mundo. (29 de 08 de 2020). *BBC News Mundo*. Obtenido de www.bbc.com: <https://www.bbc.com/mundo/noticias-53955394>
- Bishop, R. C., & beim Graben, P. (2016). Contextual Emergence of Deterministic and Stochastic Descriptions. *Atmanspacher, H., Müller-Herold, U. (eds) From Chemistry to Consciousness. Springer, Cham.*, 95-110. Obtenido de http://dx.doi.org/10.1007/978-3-319-43573-2_6
- Boshara, R., & Rademacher, I. (2021). The Future of Building Wealth. Brief Essay on the Best Ideas to Build Wealth - for everyone. En R. Devol, & D. Shideler, *Building Wealth by Investing in Four Forms of Capital* (págs. 293-300). Washington D.C.: Federal Reserve Bank of St. Louis and the Aspen Institute. Obtenido de <https://www.aspeninstitute.org/wp-content/uploads/2021/10/the-future-of-building-wealth.pdf>
- Bradshaw, M., Van de Graaf, T., & Connolly, R. (2019). Preparing for the new oil order? Saudi Arabia and Russia. *Energy Strategy Reviews*(26), 12. Obtenido de <https://doi.org/10.1016/j.esr.2019.100374>
- Brown, R. E. (2020). Donald O. Hebb and the Organization of Behavior: 17 years in the writing. *Brown Molecular Brain*, 1-28. Obtenido de <https://doi.org/10.1186/s13041-020-00567-8>
- Centro de Estudios de las Finanzas Públicas. (2020). *Coberturas petroleras 2008-2020*. Ciudad de México: Cámara de Diputados. Obtenido de <https://www.cefp.gob.mx/publicaciones/nota/2021/notacefp0132021.pdf>
- Chen, Y., He, K., & Tso, G. K. (2017). Forecasting Crude Oil Prices: a Deep Learning based Model. *International Research Journal of Engineering and Technology (IRJET)*, 300-307. Obtenido de <https://doi.org/10.1016/j.procs.2017.11.373>
- Chibani, S., & Coudert, F.-X. (2020). Machine learning approaches for the prediction of materials properties. *APL Materials*, 8(8), 10. Obtenido de <https://doi.org/10.1063/5.0018384>
- CME Group. (24 de October de 2022). *CME Group Benchmark Administration Limited*, 1.3. Obtenido de <https://www.cmegroup.com/market-data/cme-group-benchmark-administration/files/cvol-methodology.pdf>

- Colmenares, F. (Diciembre de 2008). Petróleo y Crecimiento económico en México 1938-2006. *Economía UNAM*, 53-65. Obtenido de <http://www.economia.unam.mx/publicaciones/econunam/pdfs/15/04colmenares.pdf>
- Congreso General de los Estados Unidos Mexicanos. (25 de 11 de 2020). Ley de Ingresos de la Federación. Ciudad de México, México. Obtenido de <https://www.gob.mx/senasica/documentos/ley-de-ingresos-de-la-federacion-para-el-ejercicio-fiscal-de-2022?state=published>
- Corbet, S., Hou, Y., Hu, Y., & Oxley, L. (2021). Volatility spillovers during market supply shocks: The case of negative oil prices. *Resources Policy*, 74, 11. Obtenido de <http://dx.doi.org/10.1016/j.resourpol.2021.102357>
- Daneshvar, A., Ebrahimi, M., Salahi, F., Rahmaty, M., & Homayounfar, M. (05 de 05 de 2022). Brent crude oil price forecast utilizing deep neural network architectures. *Computational Intelligence & Neuroscience*, 1-13. Obtenido de <https://doi.org/10.1155%2F2022%2F6140796>
- El-Shahat, A. (2018). Advanced Applications for Artificial Neural Networks. En L. Guesmi, H. Fathallah, & M. Menif, *Modulation Format Recognition Using Artificial Neural Networks for the Next Generation Optical Networks* (págs. 11-27). Intech. doi:10.5772/intechopen.68505
- Fotios et al, P. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3), 705-871. Obtenido de <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Galeeva, R., & Ronn, E. I. (2020). *Prices and Volatilities of Oil Markets in 2020: Back to Bachelier*. New York: Department of Finance and Risk Engineering NYU. Obtenido de <http://dx.doi.org/10.2139/ssrn.3696081>
- Gupta, V., & Pandey, A. (2018). Crude Oil Price Prediction Using LSTM Networks. *International Journal of Compture Science and Information Engineering*, 12(3). Obtenido de <https://doi.org/10.5281/zenodo.1316680>
- Gupta, N., & Nigam, S. (Abril de 2020). Crude Oil Price Prediction Using Artificial Neural Network. *Procedia Computer Science. The 3rd International Conference on Emerging Data and Industry*, 642-647. Obtenido de <http://dx.doi.org/10.1016/j.procs.2020.03.136>
- Han, S.-H., Kim, S., & Youn, Y. C. (17 de September de 2018). Artificial Neural Network: Understanding the Basic Concepts without Mathematics. *Dement Neurocogn Disord*, 83-89. Obtenido de <https://doi.org/10.12779%2Fdnd.2018.17.3.83>
- Hardesty, L. (14 de April de 2017). *MIT News on Campus and Around the World*. Obtenido de Explained: Neural newtorks. Ballyhooed artificial-intelligence technique known as "deep" learning" revives 70-year-old idea.: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414#:~:text=The%20first%20trainable%20neural%20network,between%20input%20and%20ouput%20layers>.
- Hill, T., Marquez, L., Marcus, O., & Remus, W. (1994). *Artificial Neural Network Models For Forecasting and Decision Making*. Honolulu, HI USA: University of Hawaii. Obtenido de [https://doi.org/10.1016/0169-2070\(94\)90045-0](https://doi.org/10.1016/0169-2070(94)90045-0)

- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *In advances in Neural Information Processing Systems 9*. MIT Press, Cambridge MA. Presented at NIPS 96. Obtenido de <http://dx.doi.org/10.1162/neco.1997.9.8.1735>
- Hopfield, J. J. (April de 1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, 79, 2554-2558. Obtenido de <https://doi.org/10.1073%2Fpnas.79.8.2554>
- Huang, J., Chai, J., & Cho, S. (2020). *Deep learning in finance and banking: A literature review and classification*. China: Frontiers of Business Research in China. Obtenido de <https://doi.org/10.1186/s11782-020-00082-6>
- IEA. (2022). *World Energy Outlook 2022*. Paris: License: CC BY 4.0 (report). Obtenido de <https://www.iea.org/reports/world-energy-outlook-2022>
- Institute for Economics & Peace. (2022). *Global Peace Index*. Sydney: IEP. Obtenido de <https://www.visionofhumanity.org/wp-content/uploads/2022/06/GPI-2022-web.pdf>
- Instituto Belisario Domínguez. (Diciembre de 2017). Las coberturas petroleras como medio de protección para las finanzas públicas. CDMX, Ciudad de México, México. Obtenido de <http://bibliodigitalibd.senado.gob.mx/bitstream/handle/123456789/3775/1%20Publicaci%C3%B3n%20Coberturas%20petroleras.pdf?sequence=6&isAllowed=y>
- Instituto Belisario Domínguez. (Agosto de 2022). Reporte Trimestral de los Ingresos Presupuestarios. Segundo Trimestre de 2022. 19. CDMX, Ciudad de México, México: Senado de la República. Obtenido de http://bibliodigitalibd.senado.gob.mx/bitstream/handle/123456789/5706/20220812_Reporte%20Trimestral%20de%20Ingresos%20T2022.pdf?sequence=1&isAllowed=y
- Kingma, D. P., & Ba, J. (2015). Adam. A Method for Stochastic Optimization. *3rd International Conference for Learning Representations* (pág. 15). San Diego: arxiv. Obtenido de <https://doi.org/10.48550/arXiv.1412.6980>
- Kulkarni, S., & Haidar, I. (Junio de 2009). Forecasting Model for Crude Oil Price Using Artificial Neural Network and Commodity Future Prices. *International Journal of Compture Science and Information Security*, 2(1). Obtenido de <https://doi.org/10.48550/arXiv.0906.4838>
- Kulshreshtha, S., & Vijayalakshmi, A. (25 de Junio de 2020). An Arima-LSTM Hybrid Model for Stock Market Prediction Using Live Data. *Journal of Engineering Science and Technology Review*, 13(4), 117-123. Obtenido de <https://dx.doi.org/10.25103/jestr.134.11>
- Li, Y., Harfiya, L., Purwandari, K., & Lin, Y.-D. (2020). Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model. *MDPI Sensors*, 20. Obtenido de <https://doi.org/10.3390/s20195606>
- Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, 14(29), 21. doi:10.3389/fncom.2020.00029

- Ma, R. R., Xiong, T., & Bao, Y. (2021). The Russia-Saudi Arabia oil price war during the COVID-19 pandemic. *Energy Economics*(102), 13. Obtenido de <https://doi.org/10.1016/j.eneco.2021.105517>
- Manjarrez, L. (9 de Junio de 2014). *Relaciones Neuronales Para Determinar la Atenuación del Valor de la Aceleración Máxima en Superficie de Sitios en Roca Para Zonas de Subducción*. Ciudad de México: UNAM. Obtenido de https://www.researchgate.net/publication/315762548_Relaciones_Neuronales_Para_Determinar_la_Atenuacion_del_Valor_de_la_Aceleracion_Maxima_en_Superficie_de_Sitios_en_Roca_Para_Zonas_de_Subduccion
- Matlab. (2022). *Mathworks*. Obtenido de Matlab: <https://la.mathworks.com/help/stats/fsrftest.html>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27, 12. Obtenido de <https://doi.org/10.1609/aimag.v27i4.1904>
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 115-133. Obtenido de <https://doi.org/10.1007/BF02478259>
- Meneses-Bautista, F. D., & Alvarado, M. (Mayo de 2017). Pronóstico del tipo de cambio USD/MXN con redes neuronales de retropropagación. *Research in Computing Science*, 1(113), 97-110. Obtenido de <https://doi.org/10.13053/rcs-139-1-8>
- Miller, R. G., & Sorrell, S. R. (2013). *The future of oil supply*. United Kingdom: Royal Society Publishing. Obtenido de <https://doi.org/10.1098/rsta.2013.0179>
- Minsky, M., & Papert, S. (1970). A Review of "Perceptrons: An introduction to Computational Geometry". *Information and Control*, 17(5), 501-522,. Obtenido de [https://doi.org/10.1016/S0019-9958\(70\)90409-2](https://doi.org/10.1016/S0019-9958(70)90409-2)
- Mohand, A. T., & Gaitonde, D. V. (2018). A Deep Learning Based Approach to Reduced Order Modeling for Turbulent Flow Control Using LSTM Neural Networks. *Center for Nonlinear Studies, Los Alamos National Laboratory*, 23. Obtenido de <https://doi.org/10.48550/arXiv.1804.09269>
- Muraina, I. O. (2022). Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts. *7th International Mardin Artuklu Scientific Researches Conference*, (pág. 10). Mardin, Turkey. Obtenido de https://www.researchgate.net/publication/358284895_IDEAL_DATASET_SPLITTING_RATIOS_IN_MACHINE_LEARNING_ALGORITHMS_GENERAL_CONCERNS_FOR_DATA_SCIENTISTS_AND_DATA_ANALYSTS
- Nkoro, E., & Uko, A. K. (2016). Autoregressive Distributed Lag (ARDL) cointegration technique: application and interpretation. *Journal of Statistical and Econometric Methods*, 5(4), 63-91. Obtenido de https://econpapers.repec.org/RePEc:spt:stecon:v:5:y:2016:i:4:f:5_4_3
- Oleson, M. (2004). Exploring the Relationship between Money Attitudes and Maslow's Hierarchy of Needs. *International Journal of Consumer Studies*, 83-92. Obtenido de <https://doi.org/10.1111/j.1470-6431.2004.00338.x>

- Pérez Ramirez, F., & Fernández Castaño, H. (Junio de 2007). Las redes neuronales y la evaluación del riesgo de crédito. *Rev. ing. univ. Medellín*, 6(10), 77-91. Obtenido de <https://revistas.udem.edu.co/index.php/ingenierias/article/view/225>
- Peters, B. (2018). The Computer Never Was a Brain, or the Curious Death and Designs of John von Neumann. *deutsche forschungsgemeinschaft, Phillips Universität Marburg*, 1-12. doi:10.14361/9783839442067-007
- Ramones Fernández, F. (27 de Agosto de 2014). La importancia del precio del petróleo en los ingresos petroleros. CDMX, Ciudad de México, Mexico. Obtenido de <https://ciep.mx/bkze>
- Ruiz-Porras, A., & Anguiano-Pita, J. (2021). Los precios de petróleo y la actividad económica en México. *Ensayos Revista de Economía*, 2(40), 159-188. Obtenido de <https://doi.org/10.29105/ensayos40.2-3>
- Schmidt, J., Marques, M. R., Botti, S., & Marques, M. A. (2019). Recent advances and applications of machine learning in solid-state materials science. *NPJ Computational Materials*, 1-36. Obtenido de <http://dx.doi.org/10.1038/s41524-019-0221-0>
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing*, 45(11), 9. doi:10.1109/78.650093
- Secretaría de Hacienda y Crédito Público. (2019). *Informes sobre la situación económica, las finanzas públicas y la deuda pública. Cuarto trimestre 2019*. Ciudad de México: SHCP. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2019/ivt/01inf/itindc_201904.pdf
- Secretaría de Hacienda y Crédito Público. (2010). *Informes Sobre la Situación Económica, las Finanzas Públicas y la Deuda Pública. Segundo Trimestre 2010*. CDMX, Ciudad de México, México. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2010/iit/01inf/itindc_201002.pdf
- Secretaría de Hacienda y Crédito Público. (2018). *Informes sobre la situación económica, las finanzas públicas y la deuda pública. Cuarto trimestre 2018*. Ciudad de México: SHCP. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2018/ivt/01inf/itindc_201804.pdf
- Secretaría de Hacienda y Crédito Público. (2019). *Informes sobre la situación económica, las finanzas públicas y la deuda pública. Cuarto Trimestre 2019*. Ciudad de México: SHCP. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2019/ivt/01inf/itindc_201904.pdf
- Secretaría de Hacienda y Crédito Público. (2020). *Informes sobre la situación económica, las finanzas públicas y la deuda pública. Cuatro trimestre 2020*. Ciudad de México: SHCP. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2020/ivt/01inf/itindc_202004.pdf

- Secretaría de Hacienda y Crédito Público. (2021). *Informe sobre la situación económica, las finanzas públicas y la deuda pública. Cuarto Trimestre 2021*. Ciudad de México: SHCP. Obtenido de https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2021/ivt/01inf/itindc_202104.pdf
- Shah, J., & Kiruthiga, G. (2020). Crude Oil Price Forecasting Using ARIMA Model. *International Research Journal of Engineering and Technology (IRJET)*, 7(03), 5285-5287. Obtenido de <https://www.irjet.net/archives/V7/i3/IRJET-V7I31061.pdf>
- SHCP. (2022). *Reporte de Finanzas Públicas y Deuda Pública - 4T 2021*. México: SHCP. Obtenido de https://portalhcd.diputados.gob.mx/LeyesBiblio/pdf/LIF_2021_251120.pdf
- Sivaprakash, J., & Manu, K. S. (August de 2021). Forecasting Crude Oil Price Using Artificial Neural Network Model. *Asian Journal of Management*, 12(3), 321-326. Obtenido de <http://dx.doi.org/10.52711/2321-5763.2021.00049>
- Vázquez Pérez, J. T. (21 de Abril de 2020). El papel de las coberturas petroleras en las finanzas públicas ante la caída de los precios del petróleo. CDMX, Ciudad de México, México. Obtenido de <https://ciep.mx/2YbS>
- Villada, F., Arroyave, D., & Villada, M. (2014). Pronóstico del Precio de Petróleo Mediente Redes Neuronales Artificiales. *Centro de Información Tecnológica*, 25(3), 145-154. Obtenido de <http://dx.doi.org/10.4067/S0718-07642014000300017>
- Villalón, J. G., Rodríguez Ruiz, J., & Seijas Macías, A. (2014). Evolución de los Métodos Cuantitativos Económico-Financiero-Actuarial. *XXI Jornadas ASEPUMA – IX Encuentro Internacional* (pág. 21). España: Anales de ASEPUMA. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=4749588>
- Weatherall, J. O. (2013). *Cuando los físicos asaltaron los mercados. La historia de cómo se trató de predecir lo impredecible*. Barcelona: Ariel.
- Widrow, B., & Lehr, M. A. (1993). Artificial Neural Networks of the Perceptron, Madaline, and Backpropagation Family. *Elsevier Science Publishers B.V.*, 133-205. Obtenido de <https://doi.org/10.1016/B978-0-444-89958-3.50013-9>
- Yao, L., Pu, Y., & Qiu, B. (2 de Diciembre de 2021). Prediction of Oil Price Using LSTM and Prophet. *International Conference of Applied Energy 2021*, 18. Obtenido de <https://doi.org/10.46855/energy-proceedings-9195>
- Yu, Y., Guo, S., & Chang, X. (2022). Oil prices volatility and economic performance during COVID-19 and financial crises of 2007–2008. *Resources Policy*, 75, 10. Obtenido de <https://doi.org/10.1016/j.resourpol.2021.102531>
- Zaremba, W., Sutskever, I., & Vinyals, O. (2015). Recurrent Neural Network Regularization. *The International Conference on Learning Representations* (págs. 1-8). New York: University of Cornell. Obtenido de <https://doi.org/10.48550/arXiv.1409.2329>