

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS E INGENIERÍA**



**PUCP**

**PREDICCIÓN DE PRECIOS DE COMMODITIES EMPLEANDO**

**DATA ANALYTICS Y MACHINE LEARNING**

**TRABAJO DE INVESTIGACIÓN PARA LA OBTENCIÓN DEL  
GRADO DE BACHILLER EN CIENCIAS CON MENCIÓN EN  
INGENIERÍA INDUSTRIAL**

**AUTOR**

Almudena Fosca Gamarra

**ASESOR:**

Oscar Enrique Miranda Castillo

Lima, 09 de diciembre del 2019

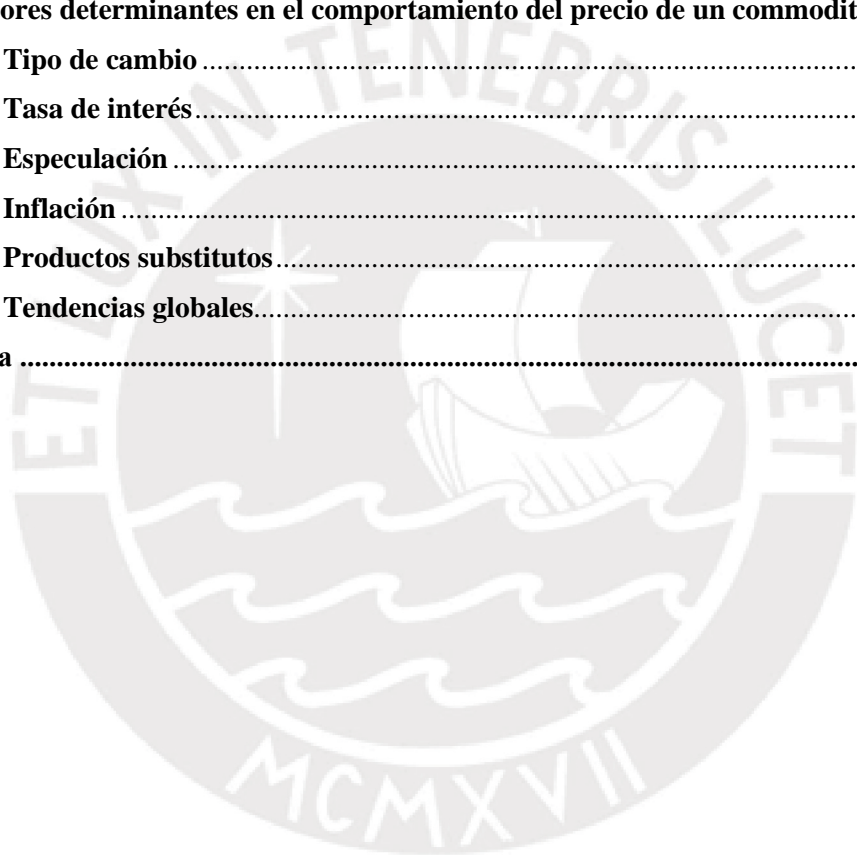
## Resumen

Esta investigación explorará las aplicaciones de herramientas de *Machine Learning* en el campo financiero. Se incluyen trabajos previos para el pronóstico de acciones, índices bursátiles y commodities, pudiendo comparar y contrastar los resultados obtenidos al aplicar diversos algoritmos. De esta forma, se emplean los estudios previos presentados como base para la elaboración de una tesis de bachillerato que tiene como objetivo pronosticar el precio del cobre empleando modelos de *Machine Learning*.



## TABLA DE CONTENIDOS

Índice de figuras.....	4
Introducción .....	5
<b>1.1 Big Data, Machine Learning y Data Analytics en la actualidad .....</b>	<b>6</b>
<b>1.1.1 Aplicaciones en el sector financiero.....</b>	<b>6</b>
<b>1.1.1.1 Precios de las acciones .....</b>	<b>8</b>
<b>1.1.1.2 Índices Bursátiles .....</b>	<b>14</b>
<b>1.1.1.3 Commodities .....</b>	<b>17</b>
<b>1.2 El Cobre .....</b>	<b>20</b>
<b>1.3 Factores determinantes en el comportamiento del precio de un commodity .....</b>	<b>24</b>
<b>1.3.1 Tipo de cambio .....</b>	<b>24</b>
<b>1.3.2 Tasa de interés.....</b>	<b>25</b>
<b>1.3.3 Especulación .....</b>	<b>26</b>
<b>1.3.4 Inflación .....</b>	<b>26</b>
<b>1.3.5 Productos sustitutos.....</b>	<b>27</b>
<b>1.3.6 Tendencias globales.....</b>	<b>27</b>
<b>Bibliografía .....</b>	<b>28</b>



## Índice de Figuras

Figura 1.1	Árbol de decisión representativo de la respuesta de un cliente al mail directo.....	6
Figura 1.2	Estructura del algoritmo Random Forest.....	7
Figura 1.3	Estructura de una red neuronal.....	9
Figura 1.4	Estructura de una red neuronal.....	10
Figura 1.5	Distribución del consumo de cobre según el tipo de mercado.....	16
Figura 1.6	Evolución del precio del cobre entre 1900 y 2015.....	17
Figura 1.7	Comparación del índice del dólar y el precio del cobre.....	18
Figura 1.8	Comparación índice real precio de los commodities (Moody's) y la tasa de interés.....	21



## Introducción

El presente trabajo de investigación surge en base al interés en desarrollar una Tesis de Bachillerato que presente el modelamiento del comportamiento del precio del cobre utilizando herramientas de *Data Analytics* y *Machine Learning*.

Se recopilaron entonces, diversas fuentes de estudios anteriores desarrollados en el campo de las finanzas y *Data Analytics*, con el fin de evaluar que herramienta es más precisa en la predicción de diversos activos financieros.



# 1. Big Data, Machine Learning y Data Analytics en la actualidad

En primer lugar, es importante hacer una distinción entre las definiciones que se abordarán a lo largo de la investigación:

En el contexto de la era de la información, Shneiderman (2008) describe un conjunto de datos como *Big Data* cuando es demasiado grande para caber en una pantalla. En pocas palabras, cuando la información no puede ser procesada de una forma directa debido a la cantidad de elementos presentes, estamos hablando de “Big Data”.

Por otro lado, “Machine Learning” es una de las ramas de los algoritmos computacionales diseñados para emular la inteligencia humana en el conocimiento y entendimiento del entorno.

Finalmente, el término *analytics* a menudo se usa ampliamente para señalar metodologías o procedimientos que permitan la toma de decisiones basada en datos. En el mundo corporativo, un equipo de análisis usa su experiencia en estadística, *big data and data mining*, *machine learning*, y *visualization* para solucionar cuestiones que se plantean sus líderes.

La interrelación de estos tres componentes junto con otros elementos de análisis y con el uso de herramientas informáticas posibilitan la evolución de la ciencia de datos.

Hoy en día la información es tal vez el activo máspreciado de las compañías, quienes la recolectan de forma “cruda” (es decir en forma de datos, la mayoría desestructurados) y procesan de acuerdo con los requerimientos de cada organización, con el fin de utilizarla para los procesos de toma de decisiones.

La inspección, limpieza, transformación y modelado de datos es el objetivo de los analistas de datos, quienes después de estudiar la data e interpretar los resultados tienen la posibilidad de brindar una recomendación o ejecutar una acción determinada.

## 1.1 Aplicaciones en el sector financiero

New Vantage publicó un informe titulado “*Big Data and AI Executive Survey 2019*” en el cual brinda una visión de cómo la ciencia de datos y la inteligencia artificial están acelerando la transformación de los negocios. En él exponen que el 97.2% de las empresas encuestadas están invirtiendo en Big Data e iniciativas de inteligencia artificial. De las 65 empresas participantes, la mayoría (el 74%) brinda servicios financieros.

Las empresas pertenecientes a este sector se han visto obligadas a implementar capacidades analíticas basadas en datos para aumentar el crecimiento, rentabilidad, optimizar procesos y reducir costos, minimizar riesgos, así como facilitar la regulación de cumplimientos.

Existen cuatro frentes en los que la inversión en “*Data Science*” puede beneficiar al sector financiero, estos son enfocados en: el cliente, las operaciones, el manejo y construcción de portafolios, y la regularización de cumplimientos.

Para los objetivos de esta investigación se desarrollará los casos de uso relacionados al enfoque operativo, como la optimización de capital y el análisis de impacto o reacción de mercado.

Gracias a una gran variedad de desarrollos tecnológicos en finanzas el beneficio de la ciencia de datos en estas áreas está siendo cada vez mayor, por ejemplo, la proliferación de plataformas de *trading* electrónico ha sido acompañada por un aumento en la disponibilidad de datos de mercado en formatos estructurados, logrando un incremento en la participación del mercado. Asimismo, la disponibilidad de acceso a data del mercado y protocolos de trading ayuda a los participantes del mercado a descubrir o pronosticar el precio y obtener una rentabilidad de las transacciones. (IOSCO, 2017, p. 12).

El caso de uso clave para el desarrollo de esta investigación es básicamente la identificación de ciertos factores o indicadores del mercado que pueden ser distinguidos como “*Trading Signals*”. El aprendizaje automático puede ayudar a las empresas a aumentar la productividad y a reducir los costos escaneando rápidamente y tomando decisiones basadas en más fuentes de información y en información oculta que puede perderse a los ojos de cualquier experto. (FSB, 2017, p.11 )

La clave entonces es poder identificar señales en la data, de la cual se puede hacer predicciones relacionadas al precio o al nivel de volatilidad sobre distintos horizontes de inversión. Queda entonces como objetivo principal identificar los factores determinantes de un “estado positivo” para invertir rentablemente.

El estudio del comportamiento de activos en el mercado financiero tiene dos bases, el análisis fundamental, bajo el cual se estudian factores que pueden afectar el valor de un activo, desde variables macroeconómicas hasta variables internas de la empresa, y el análisis técnico en el cual se usan datos de comportamiento pasado para predecir el movimiento del precio de los activos en el futuro.

Durante mucho tiempo, se creía que los cambios en los precios de las acciones no eran previsibles. La conocida hipótesis de “Random Walk” (Malkiel y Fama, 1970; Malkiel, 2003) y la Hipótesis del mercado eficiente (Jensen, 1978), establecen que un mercado es eficiente con respecto a un conjunto de información actual si es imposible hacer ganancias en este mercado. Sin embargo, a principios del siglo XXI, algunos economistas indicaron que los precios futuros de las acciones son al menos parcialmente

predecibles. Por lo tanto, se han explorado muchos algoritmos de predicción y han demostrado que el comportamiento del precio de las acciones puede predecirse. (Malkiel, 2003).

No obstante, ha habido varios estudios relacionados al modelado de algoritmos para el pronóstico de tendencias de comportamiento de acciones, bonos, índices bursátiles, entre otros.

## 1.2 Precios de las acciones

Seyed Enayatollah Alavi, Hasanali Sinaei, Elham Afsharirad (2015), desarrollaron un estudio para la predicción de la tendencia de precios de acciones usando técnicas de “Machine Learning”. En este estudio se utilizan diez años de datos del índice total de precios de acciones del banco Tejarat de Irán desde 2002 hasta 2012, el objetivo del experimento fue comparar el rendimiento y la precisión entre *Support Vector Machines*, *Random Forest* y *K nearest neighbour* (máquinas de vectores, bosques aleatorios y vecinos más cercanos). Los resultados del estudio mostraron que la mejor precisión promedio y el estimador F se obtienen del uso de *Random Forest*, seguido de *SVN* y finalmente de *KNN*. Concluyeron, además, que esto se debió a que *Random Forest* es un método de *clasificación conjunta* mientras que SVM and KNN son clasificadores simples, logrando ser una gran alternativa para un conjunto de datos no estacionarios con la presencia de valor atípicos y ruido.

Por otro lado, Muhammad Waqar, Hassan Dawood, Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar (2017), plantearon el desarrollo de un algoritmo para predecir el mercado de valores por análisis de componentes principales. La técnica de componentes principales (PCA), es muy efectiva para reducir la dimensionalidad de la data.

Esta técnica estadística se basa principalmente en el análisis de componentes principales, para enfrentar el problema presente en la recolección de información de una muestra de datos. Se sabe que para lograr encontrar representatividad en los estadísticos obtenidos se puede tender a tomar el mayor número posible de variables, sin embargo, al incrementar el número de variables también se incrementa el número de coeficientes de correlación, dificultando la visualización de relaciones entre estas. El objetivo del PCA es reducir la dimensionalidad existente en un *data set*, extrayendo toda la información perteneciente a este, pero con pocas variables no correlacionadas entre sí.

Las nuevas variables son combinaciones lineales de las anteriores y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra.

En su estudio investigaron el efecto de la aplicación de PCA en tres conjuntos de datos del mercado de valores (New York Stock Exchange, London Stock Exchange, Y Karachi Stock Exchange) y analizaron



la precisión relativa del modelo de clasificación. Los resultados del estudio demostraron que el uso de PCA logra reducir la redundancia de los datos, lo que resulta en la reducción de datos altamente correlacionados manteniendo únicamente aquellos necesarios para explicar el 99% de la varianza. Logrando obtener un modelo con mayor precisión. No obstante, para uno de los índices se encontró que la precisión disminuyó y esto puede explicarse por el hecho de que en algunos casos la aplicación de PCA da como resultado la pérdida de información crítica que podría ser importante para la clasificación y causa una disminución en la precisión del modelo.

De los resultados se puede concluir entonces que, la selección adecuada de componentes principales apropiados es vital para mejorar la utilidad de PCA.

Suryoday Basaka, Saibal Karb,c, Snehanshu Sahaa, Luckyson Khaidema, y Sudeepa Roy Deya (2019), hicieron un modelo para pronosticar la dirección de los precios de las acciones usando clasificadores de árbol. Los árboles de decisión son una herramienta de clasificación utilizada en machine learning con el objetivo de crear un modelo que predice el valor de una variable de destino en función de diversas variables de entrada. En este tipo de estructuras, las “hojas” representan etiquetas de cada clase y las “ramas” las características que se atribuyen a cada clase.

El proceso de navegación inicia desde la raíz del árbol hasta cada hoja, de acuerdo con el resultado de las pruebas a lo largo del camino.

La Figura 1.1 describe un árbol de decisión que razones por las cuales un cliente potencial responderá o no a un envío directo. Los nodos internos se representan como círculos, mientras que las hojas se denotan como triángulos. Dado este clasificador, el analista puede predecir la respuesta de un potencial cliente y comprender las características de comportamiento de toda la población de clientes potenciales con respecto al correo directo. (Rokach, Maimon. 2007)

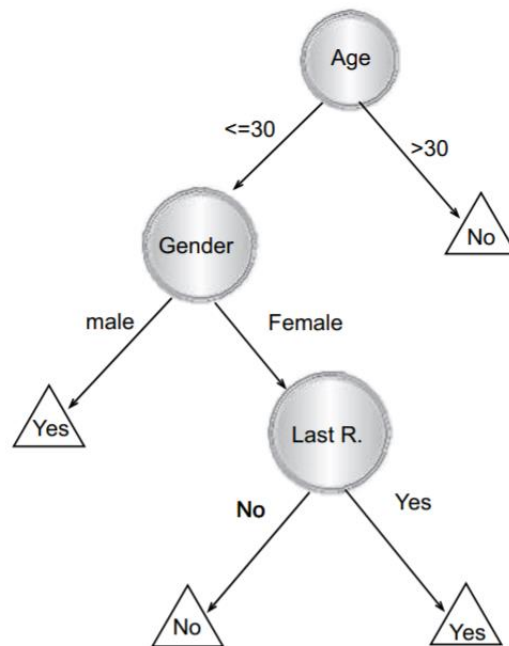


Figura 1.1: Árbol de decisión representativo de la respuesta de un cliente al mail directo  
 Elaboración: (Rokach, Maimon 2007: 2, figura 1)

Las técnicas más usadas basadas en la lógica del proceso de decisión de un árbol son métodos conjuntos híbridos entre árboles de clasificación y regresión son *gradient boosted descision trees (GBDT)* y *Random Forest (RF)*, como su nombre lo indica, consiste en una gran cantidad de árboles de decisión individuales que operan como un conjunto. Cada árbol individual en el bosque aleatorio emite una predicción de clase y la clase con más “votos” se convierte en la predicción del modelo. El uso de esta metodología permite explotar los beneficios de la lógica de un árbol de decisión pues una gran cantidad de modelos (árboles) relativamente no correlacionados que operan en conjunto superará a cualquiera de los modelos constituyentes individuales (árbol de decisión evaluado individualmente).

La razón de este efecto es que los árboles se protegen entre sí de sus errores individuales (siempre que no se equivoquen constantemente en la misma dirección). Si bien algunos árboles pueden estar equivocados, muchos otros árboles estarán en lo correcto, por lo que, como grupo, los árboles pueden predecir en la dirección correcta. (Yiu, 2019)

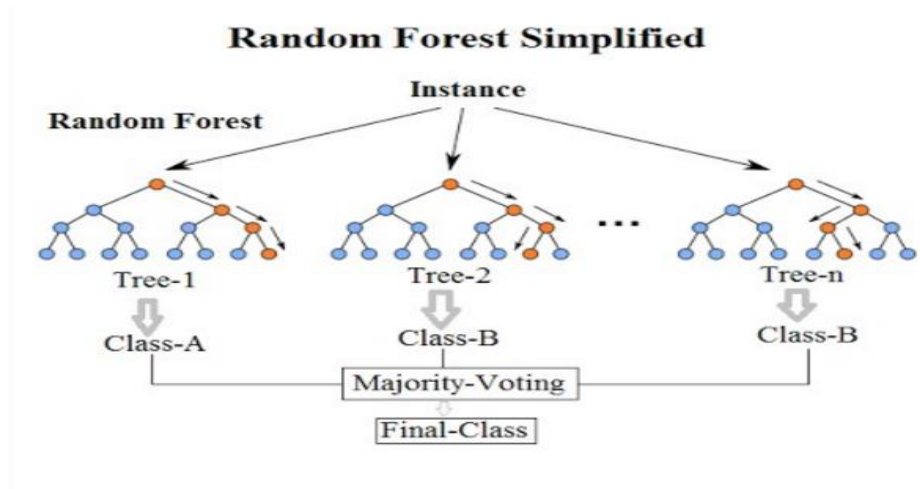


Figura 1.2: Estructura del algoritmo *Random Forest*

Elaboración: (Koheresen 2019: 1, figura 1)

Por otro lado, los modelos GBDT están enfocados en un proceso llamado “*Boosting*”, este es un método de ensamble ponderado. Cada uno de los algoritmos base se agrega secuencialmente, uno por uno. Una serie de clasificadores  $N$  aprendidos iterativamente. Los pesos se actualizan para permitir que los clasificadores posteriores “presten más atención” a las tuplas de entrenamiento que fueron clasificadas erróneamente por el clasificador anterior. *Gradient Boosting*, es conocido por ser el algoritmo líder en métodos de ensamble, utiliza el método de gradiente descendiente (algoritmo de optimización iterativa para encontrar el mínimo de una función) para optimizar la función de pérdida (o error).

La metodología del ensayo de Suryoday Basak et al (2019) fue comparar ambos algoritmos con el fin de encontrar y discutir las ventajas de estos dos sobre técnicas de análisis no ensambladas.

Las técnicas de análisis no ensambladas son llamadas también algoritmos de clasificación lineal pues utilizan técnicas de toma de decisión basadas en el valor de una combinación lineal de sus características, es decir se asume que existe un correlación lineal entre los atributos de las variables de la data.

En cuanto a los algoritmos de ensamble, son aquellos que utilizan diversos métodos base para generar un modelo óptimo, y no necesariamente asumen linealidad de datos. Como se mencionó, la aleatoriedad de los RF ayuda a hacer que el modelo sea más robusto que un solo árbol de decisión y que sea menos probable que se sobreajuste en los datos de entrenamiento. En comparación, los modelos GBDT se basan en construir un árbol a la vez, donde cada nuevo ayuda a corregir los errores cometidos por un árbol previamente entrenado logrando aproximar los regresores a las muestras de entrenamiento y encontrando la mejor división para estos.

En el experimento se utilizaron anchos de ventana (*trading window*) variante para cada corrida, entre (3,5,10,15,30,60 y 90). Este es el período en el que una compañía permite a sus ejecutivos y empleados clave transar sus acciones.

Es importante enfatizar que la diversidad de los antecedentes de las compañías elegidas para el análisis de los precios de las acciones es crucial para garantizar la eficacia de los algoritmos.

Los resultados demuestran que para ambas técnicas la precisión y el valor del estadístico F aumentan a medida que se incrementa el ancho de la ventana de trading. Más aún, la capacidad de la clasificación observada para cada caso al usar GBDT es comparable con el de RF. Por otro lado, al usar algoritmos de clasificación lineal se consiguió como máximo una precisión de 55.65%. En Conclusión, las metodologías de análisis ensamblado tienen un mejor performance en comparación a clasificadores lineales en el pronóstico del comportamiento de acciones. Además de eso, parte importante de este trabajo es la selección de indicadores técnicos y la aplicación de estos como características determinantes en el precio de las acciones. Dado que el trasfondo de esta selección es de análisis financiero, los autores recomiendan tener flexibilidad para la elección de las características, pues la interpretación de cada una depende del contexto y del mercado que se está estudiando.

Las redes neuronales artificiales (ANN) (Kimoto, Asakawa, Yoda y Takeoka, 1990; Kohara, Ishikawa, Fukuhara y Nakamura, 1997) se han empleado para lograr buenas predicciones incluso en el caso de relaciones complejas de variables.

Las redes neuronales se basan en redes de múltiples capas (nodos azules y rojos en la figura 1.3) utilizadas para clasificar variables, hacer predicciones, entre otras cosas. Existen tres partes fundamentales en una red neuronal: una capa de entrada, con unidades que representan los campos de entrada; una o varias capas ocultas; y una capa de salida. Las unidades se conectan por ponderaciones o probabilidades.

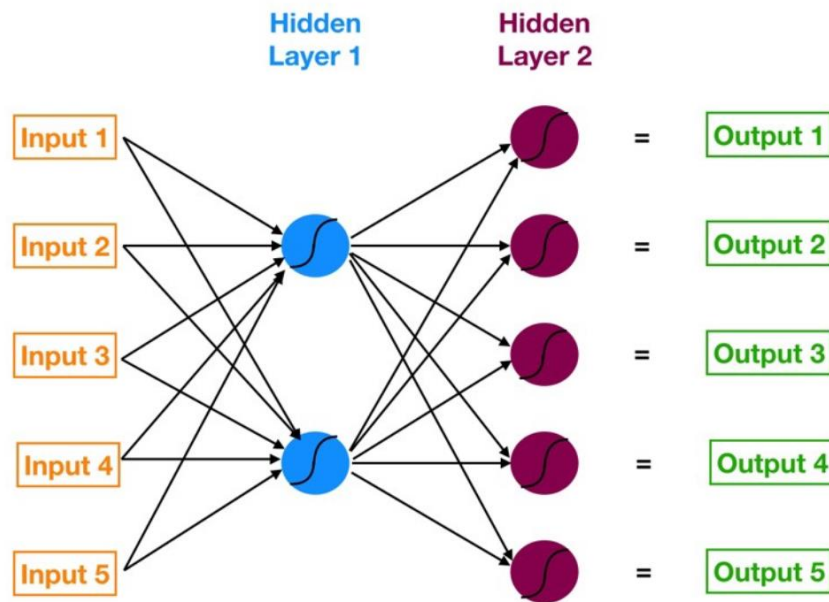


Figura 1.3: Estructura de una red neuronal  
 Elaboración: (Yiu 2019: 1, figura 2)

En un inicio, las ponderaciones son aleatorias, a medida que la red aprende examinando registros individuales, genera una predicción para cada registro y realiza ajustes a las ponderaciones iniciales dependiendo de la verdad o falsedad de cada predicción. El proceso se realiza indefinidamente hasta lograr alcanzar un criterio determinado.

Se les denomina caja negra, dado que, si bien puede abordar cualquier función, estudiar su estructura no brindará ninguna idea sobre la estructura de la función que se modela. Incluso la red neuronal más simple con una sola capa oculta es difícil de entender.

Existen redes neuronales de clases múltiples, una de ellas es la red neuronal del tipo *one-against-all* (OAA-NN). Proporciona una manera de aprovechar la clasificación binaria, en este tipo de red neuronal se clasifica bajo un clasificador binario para cada resultado posible. Se dice que este enfoque es eficiente cuando la cantidad total de clases es pequeña. Por último, las redes neuronales del tipo *one-against-one* (OAO-NN)

Boonpeng y Jeatrakul (2016) implementaron una red neuronal *one-against-all* (OAA-NN) y *one-against-one* (OAO-NN) para clasificar los datos de compra, retención o venta de acciones y compararon su desempeño con una red neuronal tradicional. Se encontró que las redes OAA-NN tenían mejor

performance que los modelos tradicionales neuronales, produciendo en promedio una exactitud de 72.5%.

### 1.3 Índices Bursátiles

Luciana S. Malagrino, Norton T. Roman, Ana M. Monteiro, (2018), llevaron a cabo el primer estudio de *machine learning* que tomaba en consideración la dependencia de índices bursátiles alrededor del mundo. El objetivo de su investigación fue pronosticar la dirección diaria del índice bursátil utilizando como metodología una Red Bayesiana (BN).

Las redes Bayesianas son una herramienta muy importante para comprender la dependencia entre los eventos y asignarles probabilidades, y con ello determinar cuán probable y cuál es el cambio de ocurrencia de un evento dado al otro. Bajo el fundamento teórico del teorema de Bayes, se utilizan probabilidades condicionales y la regla de la cadena para conseguir la distribución conjunta, es decir la probabilidad del evento final considerando todos los eventos dependientes.

En la figura 1.4 se puede visualizar un ejemplo de una red bayesiana para determinar la probabilidad de tardanza de una persona teniendo como inputs la probabilidad de situaciones como: el bus estuvo tarde y la alarma estuvo apagada.

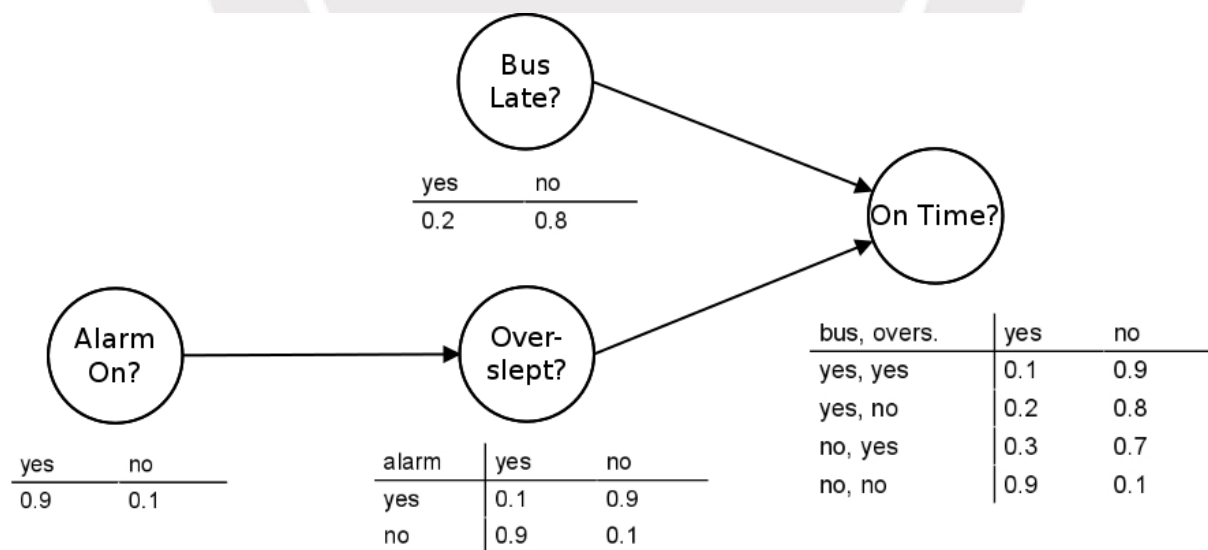


Figura 1.4: Estructura de una red neuronal  
Elaboración:(University of Bergen 2019:1, figura 1)

En el estudio realizado por Luciana S. et al (2018), se centraron en la eficiencia de las BN para lidiar con data continua y discreta, lo que las hace atractivas en la estimación del valor del precio como en el pronóstico de dirección del precio de cierre.

La hipótesis de estudio estuvo basada en que los índices se comportan bajo un enfoque “*follow the sun*”<sup>1</sup>, bajo el supuesto de que, siguiendo la secuencia de los tiempos de cierre en los mercados de todo el mundo, cada mercado cerrado influye en el siguiente.

Frente a esta hipótesis, tales redes tienen la ventaja de dar una explicación legible de las dependencias entre los mercados de valores, por lo que uno puede comprender fácilmente cuán fuertes son las dependencias al verificar la probabilidad de un cambio en el mercado dado que otro ha cambiado.

Se usaron dos metodologías, la primera tomó en cuenta las direcciones de cierre dentro de un periodo de 24 horas y otra usando un periodo de 48 horas. Agruparon los mercados bursátiles por continente e hicieron pruebas con conjuntos de 1 a 3 índices de mercado por continente. Al hacerlo, su objetivo principal era probar la viabilidad de las BN para: Identificar los mercados y, más específicamente, los índices de mercado que más influyen en iBOVESPA (índice bursátil brasileño y el elegido como objeto de estudio), y también para distinguir qué ventana de tiempo otorga el mejor análisis. Hicieron la asunción de que todos los índices dentro de un mismo continente eran independientes a la hora de la predicción, y que todos los índices de un mercado son influenciados por el índice cerrará justo antes que ellos. Es por ello que se eligen tres nodos padres sobre iBOVESPA los cuales vendrían a ser los índices bursátiles de Asia, Europa y América.

Para el experimento con 48h se sigue la misma metodología con la diferencia de que en vez de usar 3 nodos padre, se usa uno más que vendría a ser la dirección de cierre de iBOVESPA el día anterior.

Como resultado, para el caso de las 24h, se observa que la precisión media obtenida al usar un sólo índice por continente es mayor (71.08%), que las obtenidas al usar dos índices (70.20%) o tres índices (56.45%). Lo mismo ocurre con la tipología de 48h, en las que con un sólo índice se obtiene 68.17%, con dos índices 31.04% y por último con tres índices 8.21%. Frente a estos resultados, se puede deducir que cuando se añaden más índices no solo se añade más ruido a la data, sino que el modelo falla en sus predicciones a tal punto que sería mejor negar el output obtenido para que la precisión aumente.

Por otro lado, el hecho de que las 48h tuvieron peor performance que las 24h puede ser un indicador de que la reducción en la fuerza de influencia de un mercado en iBOVESPA a lo largo del tiempo se reduce, haciendo que estos mercados se comporten más como ruido conforme retrocedemos el tiempo desde el día de predicción.

---

<sup>1</sup> Según su definición básica, "seguir al sol" significa que el efecto de un índice literalmente sigue al sol: es un tipo de flujo de efecto global en el que los problemas pueden ser manejados y pasados entre diferentes zonas horarias.

En conclusión, con este modelo no solo se puede determinar las influencias en el comportamiento de un mercado sino también cuantificarlas y verificar la tasa de decrecimiento a lo largo del tiempo. Asimismo, la importancia de este es la utilidad para distinguir a los mercados candidatos a contaminarse más a causa de una crisis, en este caso los mercados más influenciados serán rankeados como más probables a pasar el efecto de la crisis en el mercado objeto de estudio. Bajo el supuesto inicial, de que el capital se mueve a los lugares donde el retorno a la inversión será mayor, se puede utilizar este modelo como herramienta de decisión para la inversión en un mercado específico.

Por último, las Redes Bayesianas son una metodología mucho más amigable y entendible para los usuarios, comparado con otros modelos tipo *black box*, como las ANNs o SVMs.

Parag C. Pendharkat, Patrick Cusatis (2018), realizaron una investigación para modelar índices financieros con *reinforcement learning agents*, un tipo de herramienta de machine learning.

El aprendizaje por refuerzo es un área de *Machine Learning*. Que trata de tomar medidas adecuadas para maximizar la recompensa en una situación particular, a diferencia del *Supervised Learning* este no se procesa con la respuesta correcta dentro del training data, sino más bien, un agente decide que acción tomar de acuerdo con los datos presentados, y en consecuencia a falta de un *training data*, aprende paulatinamente de acuerdo con el *feedback* que brinde el output del modelo.

Parag et al, se basaron en la dificultad de predecir el comportamiento financiero utilizando el modelo tradicional de toma de decisiones según la teoría de Markowitz (MDP). Dado que en la realidad ningún mercado financiero se comporta como un modelo Markovitziano, plantean abordar el problema al dividir el modelado en dos partes, en primer lugar el diseño del algoritmo basado para aprender, el modelo de la función de ganancia y las transiciones y comportamiento de los indicadores financieros, el segundo paso del MDP se encargaría de aprender el mapeo del proceso de decisiones basado en la prueba y error del modelo diseñado en la primera parte (aprendizaje reforzado).

Se aborda el diseño del algoritmo para modelar índices financieros como si fuera un problema del tipo *multi-armed bandit problem*<sup>2</sup> en el cual se desconoce la distribución de ganancia atada a cada decisión del inversor.

En este estudio se utiliza dos portafolios distintos compuestos por 2 activos financieros. El primer portafolio consistió en retornos anuales de S&P y AGG, y el segundo portafolio consistió en 3 escenarios distintos de retornos trimestrales, semi anuales, y anuales para S&P y 10 years TN. Los investigadores distinguieron a los agentes en 3: *static knowledge agent*, *continuous learning agent* y *adaptive*

---

<sup>2</sup> En la teoría de la probabilidad, los problemas llamados "multi-armed bandit problem" son aquellos en los que cada persona debe elegir entre múltiples acciones para maximizar su ganancia u obtener el resultado más rentable, teniendo en cuenta que cada acción tiene un resultado desconocido. Es un problema muy común en Casinos, al comienzo del experimento, cuando se desconocen las probabilidades y los pagos, el jugador debe determinar qué máquina tirar, en qué orden y cuántas veces. (Optipedia, 2019).



*continuous knowledge agent*. Cada uno con una función de aprendizaje distinta que se diferenciaba por la información utilizada en el proceso de estudio del training *data set*. Al analizar el impacto de la frecuencia de aprendizaje y como los agentes reaccionaron a la información brindada se obtuvo que los agentes de aprendizaje discreto o estático eran en general incapaces de utilizar todos los beneficios de una mayor frecuencia de aprendizaje, y que los agentes de aprendizaje continuo y adaptativos siempre conseguían mejor performance.

## 1.4 Commodities

Un *commodity* es un bien o producto, especialmente agrícola o minero, que puede ser procesado y revendido. Se comercializan en grandes cantidades en todo el mundo. Dependemos de ellos para las necesidades básicas de una vida cotidiana: la electricidad, alimentos, ropa, y transporte.

Antes de ahondar en el entendimiento del comportamiento de un *commodity* en el mercado financiero, es importante destacar su naturaleza física. Fundamentalmente, estos son productos creados por fuerzas naturales. Eso tiene ciertas implicaciones. En primer lugar, es que cada producto es único: su forma química depende exactamente de cuándo y dónde se originó, es decir no se pueden estandarizar en forma perfecta, como en el caso de productos manufacturados.

En general, podemos definirlos como “todo bien que es producido en masa por el hombre o del cual existen enormes cantidades disponibles en la naturaleza, que tiene valor o utilidad y un muy bajo nivel de diferenciación o especialización” (Castelo, 2003).

Para ser comercializables, los *commodities* deben ser puestos en una forma utilizable y trasladados a donde puedan ser usados, en el momento en que se necesiten. Esta relación, entre espacio, tiempo y forma, es un factor clave para entender el negocio.

Existen dos formas de clasificar los *commodities*, en primer lugar, los primarios son aquellos que se extraen o capturan de forma directa de la naturaleza, en minas, granjas y pozos, sus cualidades y características varían en gran medida. Por otro lado, los *commodities* secundarios son obtenidos a partir de los primarios para satisfacer una necesidad específica del mercado. Por ejemplo, el petróleo crudo se refina para producir gasolina y otros combustibles; Los concentrados se funden para producir metales. En este caso, puede haber variaciones menores en la calidad dependiendo de cómo se producen. Otra forma de clasificar estos productos es por su naturaleza. Los *commodities* energéticos, son los más importantes a nivel mundial, gracias a la preponderancia del petróleo. En cuanto a los productos agrícolas, las principales categorías incluyen granos y oleaginosas (maíz, soja, avena, arroz, trigo), ganado (ganado, cerdos, aves de corral), lácteos (leche, mantequilla, suero), madera, textiles (algodón, lana) y softs (cacao, café, azúcar).

Por último, la categoría de metales y minerales, incluyen metales no ferrosos, metales preciosos y minerales.

En cuanto a la comercialización de estos, se realiza en dos tipos de mercado. En primer lugar, se encuentran los mercados al contado, el cual es un término que refiere a muchos sitios descentralizados en los que el producto puede ser vendido o comprado a un precio spot acordado.

Por otro lado, el segundo tipo de mercado son los mercados listados, también llamados Bolsas de materias primas tienen una modalidad operativa que comercializa los productos mediante instrumentos derivados cuyo activo subyacente es el *commodity*. La mayoría de ellos son comercializados bajo contratos futuros, en los cuales se establece un pacto en un tiempo  $t_0$ . Por la venta de una cantidad determinada de productos ( $x$ ) a un precio  $y$  para el tiempo  $t_1$ . Las transacciones futuras representan una buena estrategia para ambas partes pues reducen el riesgo, los primeros contratos de este tipo fueron desarrollados por agricultores como una forma de reducir el riesgo de sus transacciones. Una razón para que el comercio futuro sea elegido por individuos, es la gran volatilidad que existe en el comportamiento de los precios de un *commodity*. (Yagüe, 2014).

Se han realizado investigaciones acerca modelos para la predicción de estos precios, como el realizado por Manel Hamdi Chaker Aloui, en el cual utilizan redes neuronales para predecir el precio del petróleo crudo. Se basan en la ineficiencia de estudios anteriores que utilizaron técnicas lineales para el pronóstico y obtuvieron errores significativos, dado que a pesar de haber empleado varias variables exógenas para predecir el precio del petróleo (inventario, oferta y demanda, entre otros), la oferta y demanda son relativamente inelásticas a los cambios de precios, por lo tanto, un ajuste de inventario puede ser lento, lo que explica la mayor parte de la diferencia entre los precios reales y los pronosticados, especialmente para el corto plazo (Hamilton, 2008). Sabiendo entonces que, el mercado del petróleo crudo es el mercado de materias primas más volátil, plantean que pronosticar el precio del petróleo a través de modelos no lineales es la opción adecuada.

Por otro lado, Massimo Panella, Francesco Barcellona and Rita L. D'Ecclesia (2012), especificaron en su estudio que la proyección de los precios de los productos básicos a diario no se puede obtener fácilmente utilizando modelos estructurales estándar, dada la falta de datos diarios sobre la oferta y la demanda, normalmente disponibles mensualmente. Por ello, propusieron un aprendizaje computacional de tipo ANN utilizando un enfoque de estimación de máxima precisión para calibrar los parámetros. Las redes neuronales fueron aplicadas con éxito en el modelo, logrando describir la dinámica del mercado de valores y sus volatilidades.

Como se mencionó líneas arriba, siguiendo la analogía del agricultor, al transar su cosecha a un precio determinado incluso antes de plantarla, el agricultor no debía preocuparse por los futuros cambios de

precios, sino más bien su único riesgo era la producción real de su cosecha. Por lo tanto, la compensación de un agricultor dependería de su capacidad para producir un cultivo en lugar de las posibles fluctuaciones en el precio del producto que cosecha. Al agricultor, como a cualquier otra persona que busca reducir su riesgo, se le llama *hedger*.

### **Dónde, cómo y cuándo: Los fundamentos del *pricing* de commodities**

Para entender el comportamiento de un *commodity* específico, es importante primero entender de manera general los factores macroeconómicos que intervienen en el *pricing* de estos bienes.

Como se ha enfatizado, los precios de los productos básicos tienen una gran importancia, por su potencial impacto en la producción agregada, en el equilibrio transaccional entre mercados y en la transmisión de perturbaciones del ciclo económico entre países, al conectar a los exportadores e importadores de *commodities* de países desarrollados con aquellos de los países en desarrollo. Asimismo, cambios en los niveles de precios de los productos básicos pueden crear presiones internacionales sobre la economía de un país que podrían dificultar la ejecución de las políticas monetarias en este. Si la producción de estos *commodities* constituye un porcentaje representativo del producto agregado, entonces la volatilidad de sus precios debe tenerse en cuenta en el diseño de las políticas económicas dentro de los planes de un país. (Borenzstein y Reinhart, 1994).

Identificar el comportamiento de los *commodities* permite diversificar el riesgo internacional no solo para las autoridades monetarias. Por ejemplo, si los agentes económicos en un país exportador de productos básicos supieran qué productos probablemente experimentarían aumentos de precios y el grado en que el comportamiento de estos se correlaciona pueden diversificar parte del riesgo al expandir el portafolio de productos exportadores en los que invierten actualmente. Cashin et al (1999), destaca la importancia de este análisis, pues menciona que la diversificación mediante el comercio de productos que tienen vínculos débiles y no comparten ciclos de comportamiento comunes puede disminuir sustancialmente el riesgo y el impacto de la fluctuación del precio de un único producto en la economía nacional.

Existen dos puntos de vista, según Kurgman (2008), para explicar las razones que subyacen en los movimientos de los precios de un *commodity*. En primer lugar, los niveles de vida de China y países de Asia emergente que se encuentran en veloz aumento, asimismo, poseen una alta elasticidad ingreso de la demanda de *commodities*. Dentro de este punto de vista, el valor del dólar ocupa una responsabilidad central. Por último, las condiciones monetarias laxas, y el posible exceso de liquidez internacional por

la volatilidad del dólar han ocasionado que se sumen presiones inflacionarias en mercados de precios flexibles, como en el caso de los *commodities*.

El segundo punto de vista, se enfoca en la especulación como causal de las fluctuaciones de las cotizaciones de estos productos, específicamente enfatiza la relevancia del fenómeno conocido como “financiarización de los *commodities*”, el cual se profundizará en el siguiente capítulo, esto debido a que, en la última década, se ha observado un incremento consistente del precio de los *commodities* que coincide con el volumen incrementado de transacciones e individuos en el mercado de instrumentos financieros relacionados a estos.

Si bien es cierto ambas están interrelacionadas en gran medida, se sostiene que el impacto del fenómeno mencionado, y la especulación tienen un impacto en el comportamiento a corto plazo de los *commodities*. Es importante entonces, profundizar en un estudio detallado de las variables y factores que intervienen también en el equilibrio de largo plazo de los *commodities*.

## 2. El Cobre

El cobre ha desempeñado desde siempre un papel crítico en la civilización del hombre desde tiempo prehistóricos hasta la fecha. La importancia de este metal no ferroso se basa en su alta conductividad eléctrica y térmica. Aproximadamente la mitad del cobre producido en el mundo se utiliza en aplicaciones eléctricas en distintas industrias, así como para la construcción de edificios, equipos de transporte, productos de consumo, y maquinarias y equipos industriales.

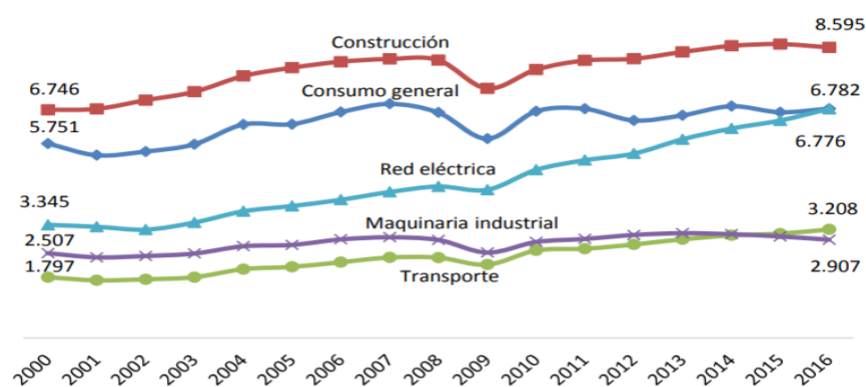


Figura 1.5: Distribución del consumo de cobre según el tipo de mercado

Elaboración: (Wood Mackenzie 2019: 1, figura 1)

En la Figura 1.5, se puede evidenciar la distribución del consumo de cobre según el mercado. Aproximadamente el 31% de la producción mundial de cobre es utilizada en construcción, siendo China el principal consumidor de cobre refinado del mundo, país en el que este metal es crucial para las metas que se ha trazado el gobierno en la última década y que actualmente están en su etapa de desarrollo. Dentro de estos planes se encuentra la modernización entera del país, electrificación del sistema nacional, disminución del carbón y petróleo incentivando energías sostenibles, implementación de proyectos de trenes rápidos, entre otros.

### El cobre en la historia

El precio del cobre ha sufrido diversos picos en la historia que vale la pena especificar, para poder encontrar los factores que han influido, y probablemente seguirán influyendo en el comportamiento de este.

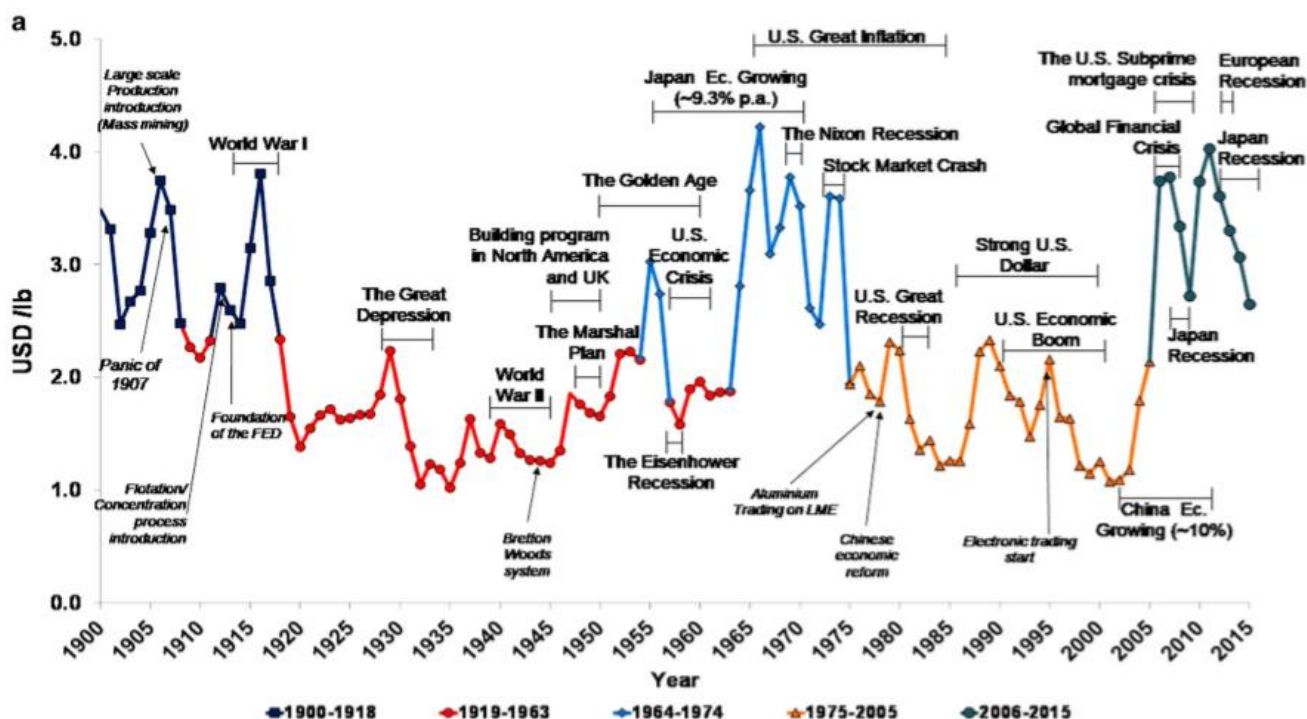


Figura 1.6: Evolución del precio del cobre entre 1900 y 2015 . Los datos incluyen eventos económicos y financieros notables en el período.

Elaboración: (Cortez et al 2018: 4, figura 1)

Vale la pena ahondar en los hechos presentados en la figura 1.6 con el fin de comprender los picos del precio en función a los hechos contextuales contemporáneos.

En primer lugar, la industria del cobre reaccionó positivamente (alza de precios) a las épocas de guerra, esto debido a que las máquinas de ataque requerían grandes cantidades de este metal para los cartuchos, balas y piezas de armas. Como se puede observar entre los años 1914-1919, el precio del cobre alcanzó el que sería su pico máximo en los siguientes 50 años.

A diferencia de la Primera Guerra Mundial, los controles de precios se impusieron desde el comienzo de la Segunda Guerra Mundial y las transacciones de cobre no se reanudaron hasta agosto de 1953.

Por otro lado, entre los años 1929 y 1935 se registró una caída del precio del cobre, coincidente con la gran depresión americana. Asimismo, si se analiza la figura detenidamente, se puede distinguir un comportamiento similar a un *momentum* entre la época dorada estadounidense y la recesión del estado americano de 1960 durante el gobierno de Eisenhower. Esta tendencia es clara también en la recesión del gobierno de Nixon, la gran recesión americana de 1980, y las crisis financieras en el mundo a partir del 2005.

En cuanto a la valorización del cobre en la historia, se identifican el periodo de crecimiento de Japón (aproximadamente 9% anual) durante el excluyendo<sup>3</sup> los años de recesión americana se registra una tasa de crecimiento anual compuesta (CAGR) de 9% para el precio del cobre. Lo mismo ocurre analizando el periodo de crecimiento de China en los inicios del 2000, obteniendo como resultado un CAGR 23% para el precio del cobre<sup>4</sup>.

Del mismo gráfico se puede observar los rastros de bajo precio están conectados por periodos de transición de alto precio que describen un momento de bajo precio "estable" que se mueve a un estado de alto precio más "volátil" durante un período corto y luego regresa al estado de bajo precio (*momentum*). De esta figura entonces, se concluye que las disminuciones en el precio global del cobre coinciden con la inflación y las recesiones en la economía más grande del mundo, Estados Unidos, y con periodos de expansión de otros países.

Si bien, se ha realizado una breve comparación entre los hechos contextuales contemporáneos y el comportamiento del precio del cobre a lo largo de los años, se debe comprender que esta es una vaga y simple forma de identificar los factores influyentes en la variabilidad de este *commodity*. Dado que los hechos no son circunstancias apartadas, sino que en su mayoría afectan al mismo tiempo y, en direcciones y magnitudes distintas al precio del cobre.

---

<sup>3</sup> Esto es, calculando el CAGR para el precio del cobre de 1965 a 1970.

<sup>4</sup> Calculado en base al 2003 como fecha inicial hasta el 2010.

En un intento por resumir los estudios sobre la dinámica de los precios de los productos básicos, Frankel y Rose (2009) enumeran tres teorías que explican el auge de estos productos en los últimos años. Las primeras dos teorías coinciden con los supuestos presentados por Kurgman.

- El primero es el "crecimiento de la demanda global", que se aceleró con la inclusión de países de alta demanda como China e India, causando los altos precios observados. Como se mencionó líneas arriba, se puede observar en la figura 1.6, que entre los años 1999 y 2010, la economía china presenció un crecimiento exponencial de aproximadamente 10% anual, el cual coincide con una tasa de crecimiento del precio del cobre.<sup>5</sup> La participación de China en las importaciones de metales aumentó de menos del 10 por ciento en 2002 al 46 por ciento en 2014.
- La segunda teoría se centra en los mercados financieros y argumenta que la "especulación" fue la causa principal del auge de los productos básicos. Dada la existencia de mercados futuros, los participantes del mercado tienen la posibilidad de mantener sus bienes a largo plazo cuando se espera que el precio del *commodity* aumente, ejerciendo una presión aún más fuerte en el precio de este.
- La última teoría se centra en las bajas tasas de interés causantes de liquidez en el mercado, desencadenando que más individuos transfieran sus inversiones de fondos a contratos de *commodities*, incrementando la demanda de estos y por lo tanto elevando el precio.

Cortez et al, (2018), investigaron el comportamiento "caótico" del precio promedio anual del cobre entre los años 1900 y 2015, con el fin de examinar la dependencia del tiempo y de atractor extraño<sup>6</sup> mediante un análisis visual de series de tiempo. Utilizaron un *dataset* de precios anual con 116 observaciones, y confirmaron que este tamaño de muestra fue adecuado para distinguir el comportamiento caótico.

Concluyeron que las series de tiempo no expresaban ningún comportamiento periódico ni eran generadas por un proceso estocástico. Además, se observó la presencia de un agente extraño que describe un momento de bajo precio "estable", interrumpido después por varios años de aumentos de precios. Llamaron a estas fluctuaciones "períodos de transición de precios" que deben abordarse como períodos de ajuste excepcionales en lugar de ciclos. Revelaron finalmente que las variables que impulsan los precios están relacionadas con el tiempo, evolucionan de una manera de causa y efecto, y los efectos se propagan con el tiempo. En su caso de estudio, se observó que las fluctuaciones de los precios de las materias primas minerales tienen efectos acumulativos a lo largo del tiempo, donde los cambios afectan no solo a los estados actuales sino también a los futuros que luego se convierten en el

---

<sup>5</sup> Ver histórico de precio de cobre en el Anexo 1

<sup>6</sup> Llámese a un factor exógeno

punto de partida para el próximo cambio de precios que describe una relación temporal continua. Este estudio es de gran valor para el entendimiento del comportamiento a largo plazo de los precios de *commodities*.

### 3. Factores determinantes en el comportamiento del precio de un commodity

En base a las teorías presentadas, y al gráfico presentado, se procede a distinguir los principales factores en la determinación del comportamiento del cobre.

#### 3.1 Tipo de cambio

Todos los trabajos de investigación utilizados en esta tesis concuerdan en el rol predominante del dólar, el cual es imposible de ignorar al estimar o modelar el comportamiento de todos los *commodities* en el mercado.

El dólar es el mecanismo de referencia universal para la fijación de precios de la mayoría de *commodities*, pues es la moneda de reserva de otros países alrededor del mundo, lo cual significa que lo mantienen como activo de reserva.

Si bien cada *commodity* tiene características idiosincráticas, históricamente, los precios de todos los productos básicos han tendido a caer cuando el dólar se fortalece frente a otras monedas, mientras que cuando el valor del dólar se debilita frente a otras monedas principales los precios generalmente tienden a aumentar.

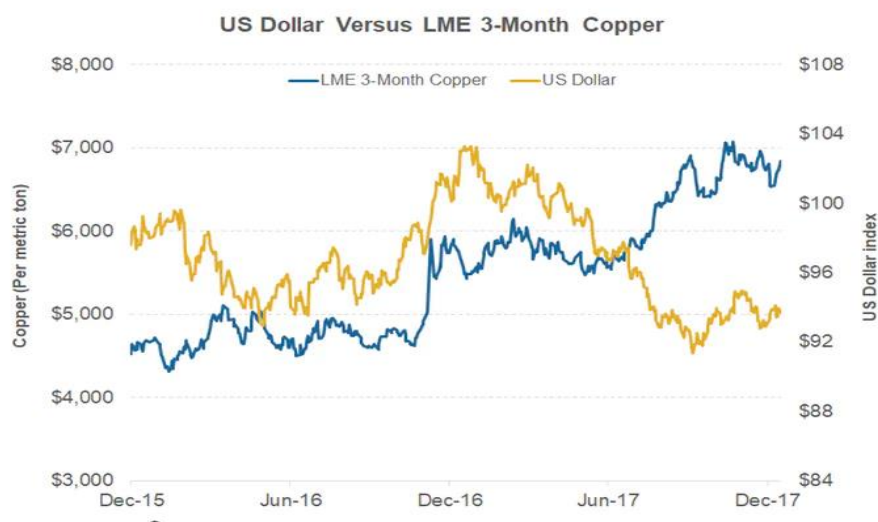


Figura 1.7: Comparación del índice del dólar y el precio del cobre

Elaboración: (LME 2018: 1, figura 1)



### 3.2 Tasa de interés

La tasa de interés es un factor que impacta directamente en el proceso de toma de decisiones de los inversionistas, lo que puede derivar en el comportamiento del precio de *commodities* a corto plazo. En un entorno de bajas tasas de interés, el costo de financiar las existencias es menor que cuando las tasas de interés son altas. Es más barato almacenar los bienes a largo plazo; el costo de inventarios es un término que los consumidores de productos básicos (y productores) usan para describir los costos asociados con la tenencia de inventarios durante un período.

Las tasas de interés están asociadas a las políticas monetarias de países desarrollados, una política monetaria “fácil”, se traduce normalmente en tasas de interés reales bajas. Barsky y Kilian (2002, 2004) han argumentado que los altos precios del petróleo y otros productos básicos en la década de 1970<sup>7</sup> no fueron exógenos, sino más bien un resultado de la facilidad de la política monetaria.

A mayor costo de mantenimiento de inventarios, el mercado responde con una baja en la demanda y por lo tanto una disminución en el precio de los *commodities*.

Asimismo, se puede decir también que en situaciones donde las tasas de interés internacionales se ubican en niveles muy bajos, los inversores buscan en otros activos financieros, como los *commodities*, alternativas más rentables, elevando la demanda de estos. (Doperto, I, Michelena, G. 2011)

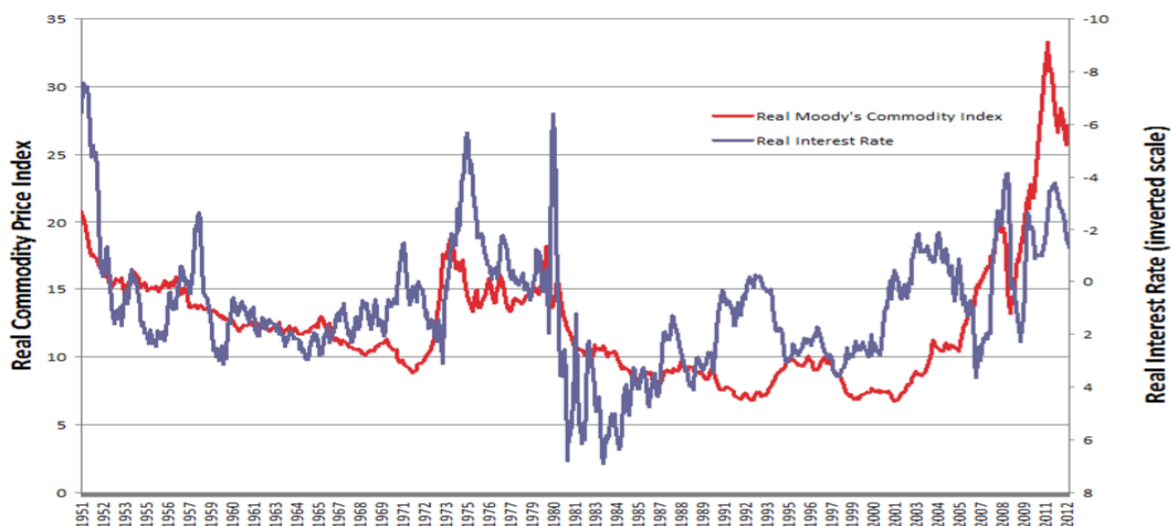


Figura 1.8: Comparación índice real precio de los commodities (Moody's) y la tasa de interés (escala invertida)

Elaboración: (Frankel, J 2014: 4, figura 1)

<sup>7</sup> Ver figura 5

### 3.3 Especulación

Podemos definir la especulación como la compra de los productos, ya sea en forma física o vía contratos negociados en una bolsa, en previsión de ganancias financieras al momento de la reventa. Esta incluye no solo la posibilidad de especulación desestabilizadora, que es la más común de considerar, sino también la posibilidad de la especulación estabilizadora (regresar el precio del *commodity* a su estado fundamental). El último caso es el fenómeno por el cual un aumento en el precio spot en relación con su equilibrio a largo plazo genera expectativas de una disminución de precios en el futuro, lo que lleva a los participantes del mercado a vender o vender en corto mercancía hoy y por lo tanto amortiguar el aumento de precios hoy. (Frankel, J 2014)

La explicación del fenómeno anterior se basa en la teoría financiera, según la cual existe especulación informada y especulación desinformada. En un mercado financiero, la primera deberá tener efectos en los precios, ya que esta es la forma en que la información privada se traduce en las transacciones del mercado. Mientras que, la especulación desinformada no debería tener tales efectos, o en casos de mercados poco líquidos no debería tener ningún efecto. Si un *trader* desinformado mueve los precios del mercado lejos de su equilibrio fundamental, *traders* informados (*arbitrageurs*), conocedores del verdadero valor, aprovecharán esta ventaja lo cual resultará en la estabilización del precio.

Si dividimos los efectos de la especulación en la transacción de *commodities* en ventajas y desventajas, obtenemos que:

En los casos en los que los productores no necesariamente desean vender a un determinado precio, o un determinado momento no coincidente con los consumidores, la presencia del *arbitrageur* hace que esta brecha sea eliminada, otorgando liquidez al mercado. Por otro lado, si no se tienen políticas reguladoras sobre el efecto de los especuladores en el precio de los *commodities*, estos pueden causar que las materias primas sean más caras para el comprador final, al estar presionando constantemente los precios.

### 3.4 Inflación

Los *commodities* representan en la actualidad activos financieros en la cartera de los inversores, por lo que los incentivos para adquirirlos como reserva de valor aumentan con el nivel de precios, es decir, con la inflación (Roache, 2010). Sin embargo, el efecto puede ser bidireccional, ya que el incremento en el precio de los *commodities* produce un incremento en el índice de precios, generando una mayor tasa de inflación.

El efecto de la inflación en el precio de los *commodities* tiene que ser estudiado detenidamente, pues si se evalúa como una relación directa con la capacidad de adquisición de los individuos participantes del mercado en cada país el efecto será independiente.

### **3.5 Productos sustitutos**

La demanda del cobre dependerá en gran medida de la existencia de posibles sustitutos. En la industria de las telecomunicaciones, el cobre es una materia prima importante, pero la promoción y aplicación de la tecnología de fibra óptica ha desafiado el estado del cobre.

El grafeno por su parte es un producto con cualidades “increíbles” según expertos, y es considerado el sustituto perfecto del cobre como conductor eléctrico y de calor. Según un estudio de Corfo, en 2006 sólo se habían hecho 138 publicaciones respecto al material en todo el mundo. En 2014, la cifra aumentó a casi 11 mil, observándose un crecimiento constante y explosivo.

### **3.6 Tendencias globales**

Los planes masivos de urbanización de China e India tendrán un fuerte impacto en la demanda del cobre. Por otro lado, existe una tendencia de “Nacionalismo de recursos” por el lado de la oferta, lo que lleva a retrasos en los proyectos y a interrupciones en el suministro. Esto se ha demostrado en varios países en desarrollo donde existen deseos del gobierno de declarar un mayor control sobre los recursos naturales ubicados en sus territorios. Cabe resaltar también la suspensión de proyectos mineros en Chile, el mayor productor mundial de cobre. Estos proyectos tenían como objetivo contribuir a una mayor capacidad productiva de cobre en 2012-2025. (Yiming W, Kabwe E, 2015).

## **Bibliografía:**

### **1. Libros:**

CASTELO M.

2003 Diccionario Comentado de Términos Financieros Ingleses de Uso Frecuente Español. A Coruña: Netbiblo.

HUTTER Frank, KOTTHOFF Lars, VANSCHOREN Joaquin

2019 Automated Machine Learning. Suiza: Springer

GÉRON Aurélien

2019 Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow. Segunda Edición. Estados Unidos: O'reilly

KUHN Max, JOHNSON Kjell

2019 Feature Engineering and Selection: A practical Approach for Predictive Models. Bookdown.org.

### **2. Página Web:**

AMAT Joaquín

2017 Máquinas de Vector Soporte. Recuperado de [https://rpubs.com/Joaquin\\_AR/267926](https://rpubs.com/Joaquin_AR/267926)

BLOOMBERG

2019 Base de datos en Bloomberg.

INSTITUTO DE INGENIEROS DE MINAS DEL PERÚ

2020 Producción de cobre en Perú creció 97% en periodo 2008-2019. Recuperado de <http://www.iimp.org.pe/actualidad/produccion-de-cobre-en-peru-crecio-97-en-periodo-2008-2019>

KOHERSEN William

2017 Understanding Random Forest. Towards Data Science. Recuperado de <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>

#### OPTIPEDIA

2019 Multi-Armed Bandit. Optimizely. Recuperado de <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>

#### SINGH Nagesh

2019 Hyperparameter Optimization for Machine Learning Models. Recuperado de <https://medium.com/swlh/hyperparameter-optimization-for-machine-learning-models-12582f00ae52>

#### SICKIT-LEARN

2019 Scikit-learn machine learning in Python. Recuperado de <https://scikit-learn.org/stable/>

#### TRAFIGURA

2018 Commodities Demystified. A guide to trading and the global supply chain. Recuperado de <https://www.bauer.uh.edu/spirrong/commoditiesdemystified-guide-en.pdf>

#### YIU Tony

2019 Understanding Random Forest: How the Algorithm Works and Why it Is So Effective. Towards Data Science. Recuperado de <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

#### YIU Tony

2019 Understanding Neural Networks. Towards Data Science. Recuperado de <https://towardsdatascience.com/understanding-neural-networks-19020b758230>

### 3. Informes y estudios:

BASTOURRE D, CARRERA J, IBARLUCIA J.

2010) Precios de los commodities: Factores estructurales, mercados financieros y dinámica no lineal. Argentina: Banco Central de la República Argentina.

#### FINANCIAL STABILITY BOARD

2017 Artificial intelligence and machine learning in financial services Market developments and financial stability implications. Recuperado de <https://www.fsb.org/wp-content/uploads/P011117.pdf>

INTERNATIONAL ORG OF SECURITIES COMMISSIONS

2017 IOSCO Research report on Fintech. *OICU-IOSCO*. Recuperado de <https://www.iosco.org/library/pubdocs/pdf/IOSCOPD554.pdf>

NEW VANTAGE PARTNERS

2019 How Big Data and AI are Accelerating Business Transformation. *Big Data and AI Executive Survey 2019*. Recuperado de <https://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf>

#### 4. Tesis, tesinas y otros similares

ALOUÏ Chaker, HAMDI Manel

2014 Forecasting Crude Oil Price Using artificial Neural Networks: A literature survey. *Economics Bulletin*. 35, 1339-1359.

BARSKY Robert, LUZT Kilian

2002 "Do We Really Know That Oil Caused the Great Stagflation? A Monetary Alternative," in *NBER Macroeconomics Annual 2001*, B. Bernanke and K. Rogoff (eds), MIT Press, Cambridge, 137-183.

BARSKY Robert, LUZT Kilian

2002 "Oil and the Macroeconomy Since the 1970s," *Journal of Economic Perspectives* 18(4), 115-134.

BORENSZTEIN Eduardo, REINHART C.M

1994 "The Macroeconomic Determinants of Commodity Prices", *IMF Staff Papers*, Vol. 41 No. 2, 236-258.

BURTON Malkiel

2003 The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, Vol. 17. No. 1. 59-82.

CASIN P, MC DERMOTT C.J, SCOTT A

1999 "The Myth of Co-Moving Commodity Prices", *Bank of New Zealand Discussion Paper*. No. G99/9.

BASAK et al

2019 Predicting the direction of stock market prices using tree-based classifiers. North American Journal of Economics and Finance, 47, pp. 552-567.

DOPORTO Miguez, MICHELENA Gabriel

2011 La volatilidad de los precios de los commodities. Comercio Exterior e Integración. 19, pp 35-53.

ENAYATOLAH et al.

2015 Predict the trend of stock prices using machine learning techniques. International Academic Journal of Economics. Vol2, No 12. pp1-11.

ELISSEFF André, GUYON Isabelle

2004 An Introduction to Variable and Feature Selection. Journal of Machine Learning Research. Vol3. pp 1-26.

FAROOQ O

2008 Commodity prices, interest rates and the dollar. Norges Bank. Oslo

FRANKEL J.,ROSE A

2009 "Determinants of Agricultural and Mineral Commodity Prices", Working Paper, Kennedy School of Government, Harvard University,

FRANKEL J.,ROSE A

2014 Effects of speculation and Interest Rates in a "Carry Trade" Model of Commodity prices. Journal of international Money and Finance. Cambridge.

FRANKLIN M, FISHER Paul Cootner, NEIL Martin

1972 An econometric model of the world copper industry. The Bell Journal of Economics and Management Science, 3, volume 2, pp 568-609.

KABWE Eugene, YIMMING Wang

2015 Analysis of copper's Market and price-focus on the last decade's change and its future trend. International Journal of scientific & technology research volume. 4, pp 1-7.

KRIECHBAUMER T, ANGUS A, PARSONS D, CASADO MR.

An improved waveletARIMA approach for forecasting metal prices. *Resour Policy* 2014;volume 39, pp 2–41.

HAMILTON J.D

2008 „,Understanding crude oil prices““ The Energy Journal, International Association for Energy Economics 30(2), 179-206.

MALAGRINO et al

2018 Forecasting stock market index daily direction: A Bayesian Network approach. *Expert Systems with Applications*. 105, pp.11-12.

MUÑOZ Ercio

2014 El efecto de sorpresas en el crecimiento de China sobre el Precio del Cobre. *Notas de investigación Journal Economía Chilena, Central Bank of Chile*. 17, pp 110-123.

PANELLA et al

2012 Forecasting Energy Commodity Prices Using Neural Networks. *Advances in Decision Sciences*. 2012. 1-26.

PENDHARKAR P, CUSATIS P

2018 Trading financial indices with reinforcement learning agents. *Expert Systems with Applications*. 103, pp.1-13

ROKACH Lior, MAIMON Oded

2007 Data Mining with Decision Trees. Tel-Aviv, Israel. World Scientific

ROACHE Shaun

2010 “What explains the rise in food price volatility?” IMF Working Paper WP/10/129.

SAVASCN Özge

2012 The Dynamics of Commodity Prices: A clustering Approach. North Carolina: University of North Carolina.

SHNEIDERMAN Ben

2008 Extreme visualization: Squeezing a billion data points into a million pixels. *ACM SIGMOD International Conference on Management of Data*. ACM, New York. Recuperado de <http://www.cs.umd.edu/~ben/papers/Shneiderman2008Extreme.pdf>



YAGÜE Pablo

2014 Estudio de los commodities: El caso de los cereales. Madrid, España: ICADE.

