ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE POSTGRADOS

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN LOGISTICA Y TRANSPORTE CON MENCION EN MODELOS DE OPTIMIZACION”

TEMA:

“PRONOSTICO DEL PRECIO DEL ORO POR MEDIO DE MODELOS AUTOREGRESIVOS INTEGRADOS DE PROMEDIO MOVIL, MODELOS RELACIONALES Y CONCEPTOS DE MACHINE LEARNING”

AUTOR:

MARIO ERNESTO ORDOÑEZ MEJIA

Guayaquil - Ecuador

2020

Resumen

La cantidad da data generada y almacenada en estos tiempos, nos pone ante el desafío de poder realizar un manejo de la información pudiendo modelizar tendencias y patrones para poder tomar decisiones más acertadas en cada uno de los campos en los que podamos desenvolvernos. Hemos tomado como caso de ejemplo el desarrollo del precio del oro a través de la historia teniendo una base de datos registrada desde finales del año 1968 con registros diarios de acuerdo con los horarios de apertura de los mercados internacionales. Esta data fue recopilada, luego ordenada, agrupándola en datos promedio mensuales y obtener información tendencial de precios y probablemente tener un análisis de trabajos posteriores más realista. Estos datos fueron trabajados usando tres tipos de metodología clásica como son los modelos autorregresivos integrados de promedios móviles (ARIMA), un modelo relacional considerando variables macroeconómicas externas y establecer las relaciones con el precio del oro y por último se ejemplificará un modelo de redes neuronales artificiales, LSTM. La cantidad de data disponible es de gran importancia para la selectividad de la metodología a usarse, fue posible constatar que un modelo ARIMA presento las medidas de error más bajas que las obtenidas por las redes neuronales (AI) esto debido a que las redes neuronales están diseñadas para trabajar con grandes cantidades de data, las redes neuronales tiene factores que pueden ser modificados e influyen en el error medido del modelo y por ultimo los tiempos y capacidades computacionales que estos métodos necesitan son mucho más altas que las que normalmente uno tiene en su hogar por medio de su computador , es decir, existen equipos computacionales diseñados especialmente para el manejo de altas cantidades de data y realización de modelos de machine learning o redes neuronales artificiales.

Palabras claves: Arima, Machine learning, redes neuronales, LSTM.

ABSTRACT

The amount of data generated and stored in these times, puts us before the challenge of being able to manage the information, being able to model trends and patterns in order to make more accurate decisions in each of the fields in which we can operate. We have taken as an example the development of the price of gold throughout history having a database registered since the end of 1968 with daily records according to the opening hours of international markets. This information was collected, then ordered, grouped into monthly average data and obtaining price trend information, and you probably have a more realistic analysis of subsequent work. These data were worked using three types of classical methodology such as the integrated autoregressive moving average models (ARIMA), a relational model relative to macroeconomic variables and establishing the relationships with the price of gold, and finally, an artificial neural network model (LSTM) will be exemplified. The amount of data available is of great importance for the selectivity of the methodology to be affected, it was possible to verify that an ARIMA model presents lower error measures than those obtained by neural networks (AI), this because neural networks are needed To work with large amounts of data, neural networks have factors that can be modified and influence the measured error of the model, and lastly, the computational times and capacities that these methods need are much higher than what one normally has at home. By means of your computer, that is, there are specifically specific computer equipment for handling high amounts of data and performing machine learning models or artificial neural networks.

Key words: Arima, Machine learning, neural networks, LSTM.

DEDICATORIA

Dedico este trabajo a Rosario, Macu, Benjo, Fiorella, Matías y Rebecca

AGRADECIMIENTO

Agradezco a Dios y Rosario, mi madre, que hicieron esto posible; Macu y Benjamín por su paciencia y apoyo en el trayecto.

A Sandra García, mi tutora, por su ayuda en la elaboración de este proyecto.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Graduación me corresponde exclusivamente; el patrimonio intelectual del mismo corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas, Departamento de Postgrados** de la Escuela Superior Politécnica del Litoral.

Mario Ernesto Ordóñez Mejia

TRIBUNAL DE GRADUACIÓN

Ph.D. María Nela Pastuizaca Fernández
PRESIDENTE

Ph.D. Sandra García Bustos
DIRECTOR

Ph.D. Omar Ruiz Barzola
VOCAL 1

Ph.D. Holger Cevallos Valdiviezo
VOCAL 2

AUTOR DEL PROYECTO

Mario Ernesto Ordoñez Mejia

Contenido

INDICE DE FIGURAS

INDICE DE TABLAS

PRESENTACIÓN

Desde tiempos antiguos los metales preciosos tales como la plata y el oro fueron objetos de valor debido a sus características físicas-químicas, usados como base o unidad transaccional de bienes y/o servicios.

Al pasar del tiempo el oro fue considerado como soporte de valor a la moneda de un país, lo que conllevó a la creación de reservas de oro en cada uno de los países emisores de esta moneda, por ejemplo, el dólar posee un soporte de valor mediante las reservas de oro de los Estados Unidos de Norteamérica. Los usos del oro han ido evolucionando en diferentes áreas tales como la joyería, tecnología e inversiones, por lo que ha habido una tendencia creciente de la demanda de este metal. La extracción del oro de la tierra ha ido teniendo en el transcurso del tiempo una evolución y en algunas ocasiones más complicadas que otras. Diferentes tipos de yacimientos y el desarrollo de nuevas metodologías, tecnologías, químicas y físicas. Todos estos factores en adición a los cambios tanto en oferta como en demanda han hecho que el desarrollo de modelos predictivos del precio de los metales y commodities, sea un campo interesante de estudio.

Sin duda alguna los avances en los procesos computacionales, así como de metodologías y desarrollo de algoritmos de machine learning, nos dan nuevos medios de análisis y resultados, los cuales son una fuente de referencia mucho más confiable y dinámica tanto para la empresa en su planificación de flujos financieros, operaciones y supply chain, así como para el inversionista en sus transacciones.

CAPÍTULO 1

INTRODUCCION

El Oro es considerado un metal precioso, de características físicas y químicas de gran importancia y su valor se relaciona con el poder económico, antiguamente era de un uso exclusivo de las realezas, ahora tiene una gran demanda para la joyería, para la industria, las aplicaciones tecnológicas y la electrónica. Este metal, se encuentra en la naturaleza en diferentes formas de depósito y con estos diferentes grados de dificultad operacional, procesos de extracción química, mecánica, combinados y esto afecta directamente a los grados de inversión para la extracción de este metal, siendo este el objetivo de la industria minera.

Los precios de los metales preciosos presentan una variabilidad a lo largo del tiempo con alguna relación a factores y/o índices macroeconómicos tales como: la inflación, tasas de interés, precio del petróleo, índice de mercado de valores, índice de la moneda dólar, de los cuales sería necesario realizar los análisis relacionales para nuestro ejercicio, como ejemplo la relación entre el precio del oro y los mercados accionarios es generalmente inversa, es decir, los inversionistas mueven el dinero del oro hacia las acciones en tiempos de boom y viceversa en tiempos de crisis; la relación entre el precio del oro y el precio del petróleo es generalmente positiva, es decir, en algún momento de tensión económica relacionada puede elevar ambos precios del petróleo y del oro (Banhi Guha & Gautam Bandyopadhyay, 2016).

El oro presenta un atractivo a toda clase de personas como propósito de inversión, las personas que invierten en oro tienen dos objetivos primordiales, primero, como una protección contra la inflación en un periodo de tiempo, y mejorar el retorno de las inversiones, segundo como herramienta diversificadora, diversificar el riesgo y ayudar a reducir la volatilidad de todo el portafolio de inversiones. Las formas de inversión en oro han tenido cambios y dan opciones distantes a los tiempos de antes con sus maneras tradicionales como comprando joyas o por las formas modernas comprando monedas y barras de oro o

invirtiendo en Exchange Traded Funds (ETF), los ETF son fondos de inversión que cotizan en la Bolsa (Banhi Guha & Gautam Bandyopadhyay, 2016).

El oro puede ser obtenido por operaciones mineras subterráneas, así como superficiales, en forma liberada o como parte de mineralizaciones de las rocas donde habita y cada modalidad representa un método diferente de obtención de oro como producto final, siendo la unidad de valor para la producción y proyecciones operativas - financieras lo que se llama la Ley de mineral, esto es, para el oro, cantidad estimada de gramos de oro por cada tonelada de material explotado, gr/Ton. Este valor permite a la empresa determinar si el proyecto es económica y operativamente viable, sumado a esto tenemos que considerar en nuestra ecuación el precio del oro actual y el precio proyectado para las planificaciones futuras. En la industria minera es difícil saber de manera exacta la cantidad de gramos de oro por cada tonelada de material explotado que hay en cierto depósito, más, es posible por métodos geoestadísticos y campañas de exploración categorizar los depósitos en Probados, Probables y Posibles siendo los primeros los de mayor grado de certeza y confiabilidad; esto por la sencilla razón que no es posible ver el interior de la tierra. Esto hace que esta industria tenga un considerable índice de riesgo para el inversionista, pero de igual manera un rendimiento más alto en comparación a otro tipo de inversiones. El poder ajustar modelos estadísticos de series de tiempo mediante métodos computacionales como con los modelos auto regresivos ARIMA y en el campo del Machine Leanring con algoritmos supervisados de regresión (OLSR) para el modelo regresivo multivariable y redes neuronales artificiales (ANN, LSTM) como método alternativo al tratamiento de series de tiempo; dará sin duda alguna una mejor proyección de los proyectos mineros, mejor planificación de operaciones y de estados financieros, una mejor capacidad de decisión para el empresario y una mejor capacidad de decisión para el inversor (Hossein Mombeini & Abdolreza yazdani, 2015).

1.1. Antecedentes y justificación

Acorde al trabajo realizado por Shahriar Shafiee y Erkan Topal, (2009) durante los años transcurridos entre enero de 2004 y diciembre de 2014, los precios de la mayoría de los commodities experimentaron variaciones muy importantes.

Desde comienzos del siglo y hasta el año 2007 ha tenido lugar un crecimiento importante en el mismo, impulsado por la fuerte demanda de materias primas generada como consecuencia de la elevada tasa de crecimiento de la economía china y de otros países emergentes, junto con la estabilidad macroeconómica global que permitió, un crecimiento consistente, aunque lento, de la mayoría de los países industrializados.

Sin embargo, ante la turbulencia provocada por la crisis financiera entre el 2007 y el 2009 cambió radicalmente ese escenario. Los agentes económicos buscaron un refugio seguro, tradicionalmente en los metales preciosos y las monedas fuertes y así se dio un impulso importante al alza del precio del oro (y, en menor medida la plata), junto con la apreciación del dólar americano frente al resto de las monedas.

Si bien durante los primeros años del siglo la tendencia en los precios de los metales fue consistentemente al alza, con una relación positiva entre ellos, solamente el oro y la plata se consideran un refugio seguro ante la incertidumbre. Por lo tanto, como consecuencia de las graves repercusiones del colapso financiero mundial sobre la actividad económica, los precios de los metales de uso principalmente industrial siguieron un comportamiento claramente diferente en relación con los metales preciosos. No obstante, la correlación entre los precios de los metales, tales como el Oro, Plata, Cobre, Acero, Aluminio, Zinc, fue muy alta durante el periodo en mención.

Según se observa en los datos históricos, Investing, (2019), el oro aumentó desde un precio cercano a los 400 dólares por onza en enero del 2004, hasta un máximo de 1,896 dólares por onza en septiembre del 2011, es decir, un crecimiento porcentual del 343%, para después registrar un ajuste importante respecto a su nivel máximo y buscar un nuevo piso alrededor de los 1200 dólares hacia el final del periodo mencionado. Por su parte, el precio de la plata pasó de poco más de 7 dólares la onza hasta casi 38 dólares entre enero del 2004 hasta marzo del 2011, es decir, un crecimiento de 442%. Desde ese momento se produjo un patrón de ajuste errático para finalmente iniciar una tendencia descendente pronunciada en septiembre del 2012 y alcanzar un mínimo nivel de poco más de quince dólares a finales del 2014. Aunque hasta el 2012 el precio

de la plata tuvo una fuerte correlación con el precio del oro, después de noviembre del 2012 parece haber disminuido, aunque la tendencia general ha sido bastante similar. En su conjunto, los minerales que sirven como insumo en procesos industriales y el acero observan un comportamiento influido principalmente por el ciclo económico, aunque en cada caso existen especificidades y diferencias por el distinto grado de concentración industrial, el cambio tecnológico, las condiciones particulares de diferentes regiones productivas.

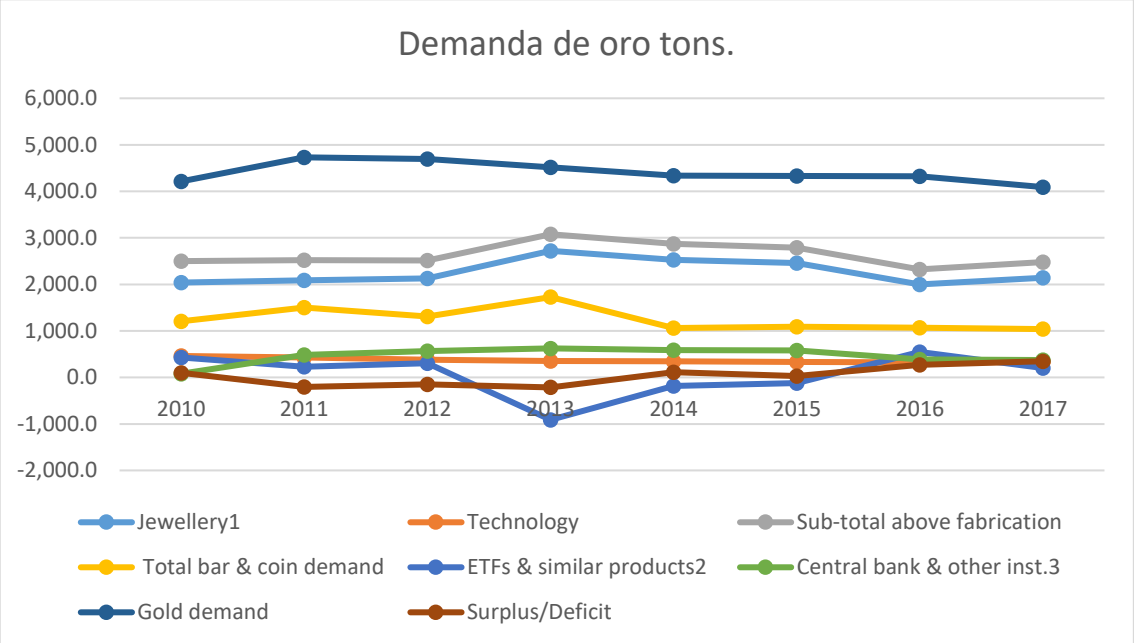
En condiciones normales, los cambios tecnológicos en la minería son muy graduales y , por lo tanto, los costos de operación de las empresas de la industria son muy estables en el tiempo, es posible suponer que las fluctuaciones de los precios de los minerales, sujetos a frecuentes altibajos, relacionados generalmente con el ciclo económico, pero en ocasiones también a situaciones geopolíticas e incluso decisiones estratégicas de los grandes productores mundiales, son un factor muy importante en la determinación de los niveles de rentabilidad y la generación de flujo operativo de las empresas mineras. Cuando los precios de los minerales son elevados, la mayor utilidad y generación de flujo operativo reducen los requerimientos de financiamiento externo, y su efecto debería verse reflejado en la estructura de capital de la empresa minera.

Los flujos de caja en proyectos mineros son volátiles y son significativamente influenciados por la fluctuación del precio de los metales. La estimación del precio de los minerales es vital en el comienzo del proceso de valoración, así como también en el cálculo de costos totales y de los índices de producción durante toda la vida de la mina, en base a esto las compañías mineras pueden aceptar o rechazar un proyecto basado en las expectativas del precio futuro. Consecuentemente es esencial estimar precios futuros con modelos adecuados durante la evaluación de proyectos mineros.

En la figura# 1 y 2 se puede observar la oferta y demanda mundial para el oro en los periodos desde el 2002 al 2007. Como se puede observar en la tabla # 2 la oferta total de oro mundial está en alrededor de 3500 toneladas por año. La fuente que más genera la oferta de oro, en aproximadamente 2500 toneladas, es la producción minera. La segunda fuente generadora de oferta de oro, son

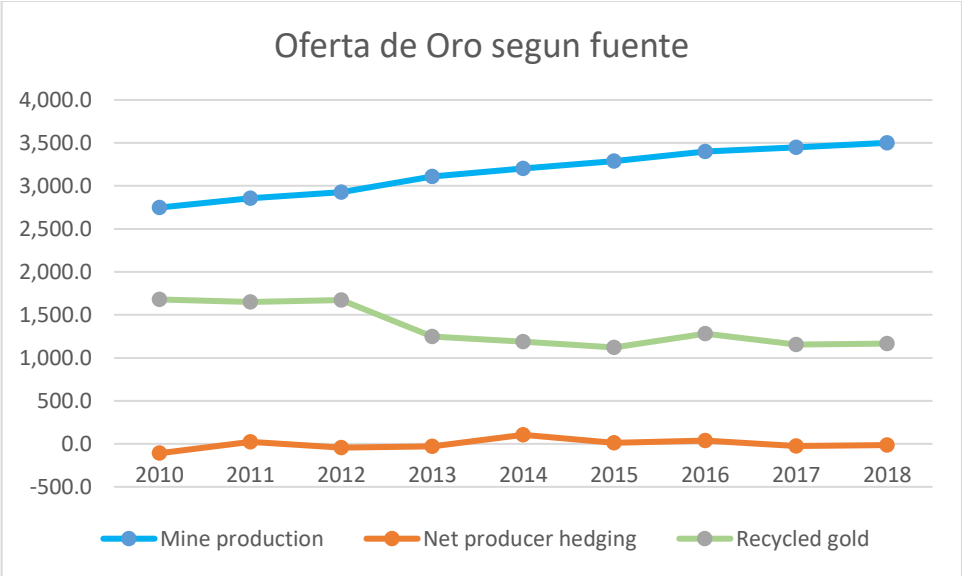
las ventas del banco central y otras. En la demanda promedio 2500 toneladas son atribuidas a la joyería y unas 1000 toneladas son atribuidas a inversionistas menores, Exchange Traded Funds (ETF´s) y la producción industrial en los últimos 10 años. La demanda mundial del oro ha decrecido, mientras que las inversiones, ETF´s y la demanda industrial se ha incrementado (World Gold Council, 2018). Una de las peculiaridades de la demanda del oro por parte del segmento de la joyería es que esta puede pasar a ser o convertirse en suministrador u ofertador. Esto significa que el oro es un recurso "renovable", sin degradación en calidad, el cual podría ser reciclado y contribuir a la disminución de la demanda global de nuevas minas de oro. En otras palabras, las reservas de oro en los bancos centrales y la joyería podrían entrar en el lado de los proveedores en la ecuación del mercado de oro.

Figura 1.1.1. Demanda mundial de oro en toneladas desde el año 2010 hasta el 2017



Fuente: Autor, estadísticas del WGC, World gold council, 2018.

Figura 1.1.2. Tipos de ofertas de oro en Toneladas



Fuente: WGC, World gold council, 2018.

En el detalle de la oferta global de oro, los mayores países productores en el 2016 China (15%), Australia (9%), Rusia (9%), Estados Unidos (7%), Perú (5%) y Sudáfrica (5%) producen alrededor del 50% del oro global en el 2016. Además, el tiempo de agotamiento o la proporción de producción minera a reservas muestra que, en promedio, las reservas mundiales de oro disminuirán en menos de 40 años, los nuevos yacimientos, reservas y producciones probablemente pospondrán esta disminución y hay que considerar que el precio del oro es uno de los factores que afectan la disponibilidad de estos yacimientos (World Gold Council, 2018).

Tendencia Histórica del precio del oro.

Estudios previos muestran que las fluctuaciones del precio del oro tienen diferentes efectos sobre la producción de oro y el valor de las acciones de las minas de oro de país a país y de mina a mina. Desde 1833 a 1933 los precios del oro fueron prácticamente constantes con un precio de alrededor 20 USD la onza y desde 1934 a 1967 hubo un incremento a 35 USD la onza después que el presidente Roosevelt fijó el precio del oro en 1934, permaneciendo estable hasta 1967 cuando el precio del oro fue liberado. El oro fue negociado en el

mercado desde 1967 y el precio se incrementó con rápidas fluctuaciones desde ese momento.

La figura 1.1.3 describe dos saltos significativos del precio del oro en la tendencia histórica. El primero fue temprano aproximadamente en enero de 1980, cuando el precio del oro superó los 600 USD y luego una pendiente negativa pronunciada hasta el año 1985, bajando el precio hasta alrededor de los 300 USD. El segundo salto comenzó aproximadamente en el 2008 con una pendiente positiva llegando a un máximo de 1895 USD en septiembre del 2011 y aparentemente en la actualidad en un estado de ajuste (Investing, 2018). Existen algunos factores que contribuyen a las escaladas del precio del oro en el corto y largo plazo.

Figura 1.1.3. Evolución histórica del precio internacional del oro

Fuente: El autor, base de datos inversiones investing.com, 2020

En el corto plazo hay dos momentos donde el precio del oro se incrementa de manera drástica. Un primer momento es el periodo donde los mercados financieros globales colapsan y una economía global en recesión, esto causa que los inversionistas confíen poco en los mercados financieros como inversiones seguras. Consecuentemente, estos inversionistas comienzan a especular o buscan otros mercados que no tengan una responsabilidad pesada, riesgo pesado o impredecibles, tal como el mercado del oro. En otras palabras, el mercado del oro opera como un tipo de seguro en contra de los movimientos extremos en los valores de los activos tradicionales durante las etapas de

inestabilidad de los mercados financieros. Segundo, la devaluación del US. dólar versus otras divisas y la inflación internacional con los altos precios del petróleo explican porque grandes compañías protegen el oro contra las fluctuaciones del US. dólar y la inflación. Esto significa que las transacciones de oro compensan el potencial movimiento del valor real en el mercado en el corto plazo contra las oscilaciones del US. dólar y la inflación.

En el largo plazo, existen dos grandes razones para el incremento del precio del oro, la oferta y la demanda del oro.

Una de las razones por las que podría haber una reducción de la oferta de oro es mediante la reducción de la producción minera y esta a su vez se disminuida por los incrementos en los costos de operación, esto decrece la exploración y lógicamente encontrar nuevos depósitos. Hablando respecto a la demanda, una de las formas en que esta se ve aumentada es cuando los inversionistas mantienen el oro en sus portafolios o la facilidad de inversión en ETF's de oro, adicionando los tiempos de incertidumbre financiera externa, producen un aumento de la demanda del metal. (Shahriar Shafiee y Erkan Topal, 2009).

1.1.1 Proyectos mineros en el Ecuador y negocios del oro.

Acorde a las estadísticas del Ministerio de Minería del Ecuador (2018) desde el 2009 aproximadamente surgieron voces de muchos proyectos que lanzaban ciertas luces de ser proyectos rentables con altas concentraciones de mineral, estas voces eran de compañías principalmente extranjeras realizando sus fases de exploración respectiva, donde resultados previos de esta labor arrojan datos probables de cantidad de mineral de interés económico. Al transcurrir de los años el escenario político del Ecuador fue cambiando de tal manera que las políticas de inversión minera cambiaron y con esto los cálculos, proyecciones de flujo y rentabilidad de los proyectos tuvieron que ser recalculados de acuerdo con estos nuevos escenarios. Ahora en el 2017 después de cambiar algunos de los protagonistas anteriores, se tiene proyectos definidos con fechas ya estimadas de inicio de producción y otros proyectos con etapas de exploración avanzada, donde se determinan la rentabilidad más segura.

Entre los principales proyectos mineros tenemos:

Proyecto Mirador, ubicado en Tundayme, Zamora Chinchipe, tienen reservas categorizadas como probadas de 137 millones de libras de cobre, 3.22 millones de onzas con una vida útil del proyecto de 30 años, actualmente están en una fase de construcción de infraestructuras preproducción.

Proyecto Fruta del Norte, ubicado en Yantzaza, Zamora Chinchipe, tiene reservas categorizadas como probadas de 4.9 millones de onzas de oro, con una vida útil proyectada de 15 años.

Proyecto Loma Larga, ubicado en Cuenca, Azuay, tiene reservas categorizadas como probadas de 1.86 millones de onzas de oro y 33.38 toneladas de cobre, con una vida útil proyectada de 11 años.

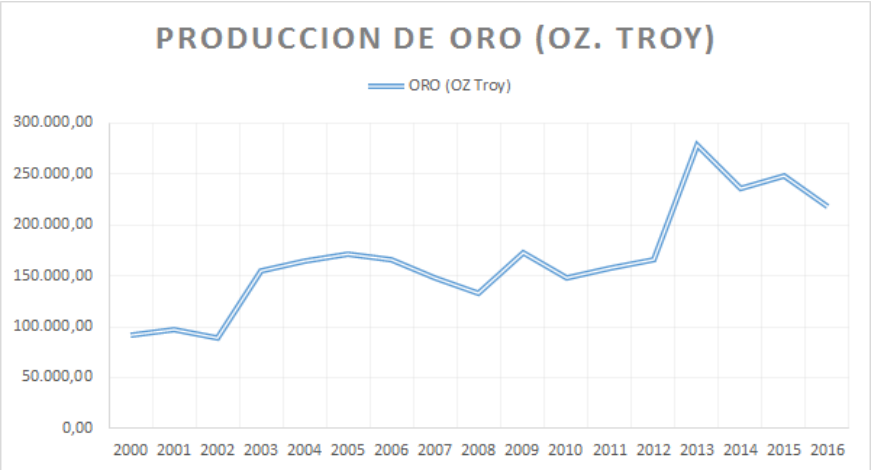
Proyecto Río Blanco, ubicado en Molleturo, Cuenca, Azuay, tiene reservas categorizadas como probadas de 605000 onzas de oro, con una vida útil proyectada de 11 años.

A estos proyectos se podrían añadir otros proyectos, los cuales , están en una etapa de exploración , es decir, solo existen inferencias de las reservas del mineral de interés, así como no está determinada la rentabilidad del proyecto, como ejemplo está el proyecto LLurimagua, ubicado en la provincia de Imbabura, a cargo de un consorcio entre ENAMI EP y la estatal chilena CODELCO, ellos están en una fase de exploración avanzada determinando las reservas y posterior rentabilidad del proyecto (Ministerio de Minería del Ecuador, 2018).

Cada uno de los minerales de interés están regidos a precios internacionales, cuyos precios están en función de algunos factores económicos. Nuestro mineral metálico de interés en esta ocasión es el oro, sumando las reservas de los proyectos mencionados tenemos una producción de 10.585 millones de onzas de oro más, ahora, esta cantidad total de producción de oro puede ser mensualizada e incluso anualizada sin problema alguno en virtud de la vida útil de los proyectos. A pesar de la proyección de producción del mineral de interés la minería del oro en el Ecuador ha tenido su desarrollo en esta industria desde hace muchos años atrás, el desarrollo técnico y tecnológico ha sido lento; existen

registros de la producción aproximada de oro en oz troy desde el año 2000 en la figura 1.1.1.1.

Figura 1.1.1.1. Producción de oro en el Ecuador desde el año 2000 hasta el 2016

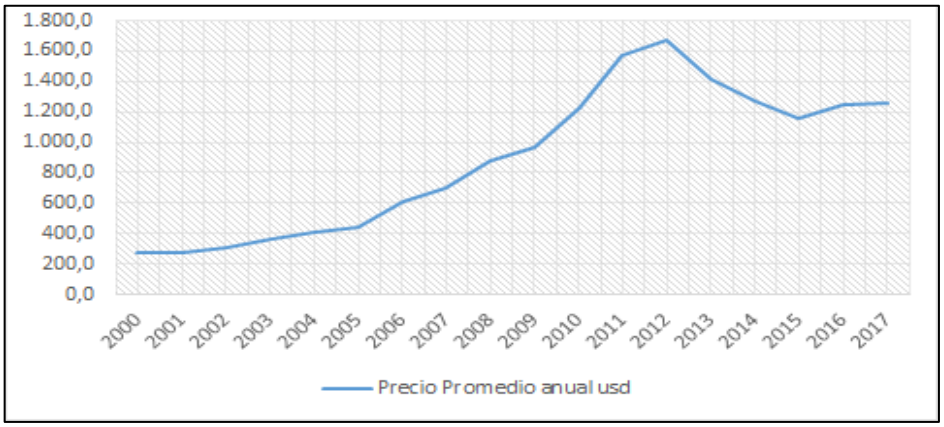


Fuente: Estadísticas del Ministerio de Minería del Ecuador, 2019.

Es posible ver la evolución histórica de la producción de oro en el Ecuador hasta el año 2016 hasta donde se tienen registros (Ministerio de Minería del ecuador, 2018). La minería en el Ecuador ha oscilado entre los regímenes de minería artesanal e informal, la minería a gran escala es un hito en la historia minera del Ecuador.

Es importante considerar la temporalidad de los niveles de producción minera para el estudio de viabilidad de los proyectos ya que, esta temporalidad actúa también sobre los precios de los metales y esto sumado a los niveles de producción, inversiones, planificación de operaciones se verán afectadas en función del precio de los metales. En la tabla abajo veremos la evolución del precio del oro en promedio anual desde el 2000 hasta 2017, base de datos Investing, (2018).

Figura 1.1.1.2. Precio promedio anual del oro



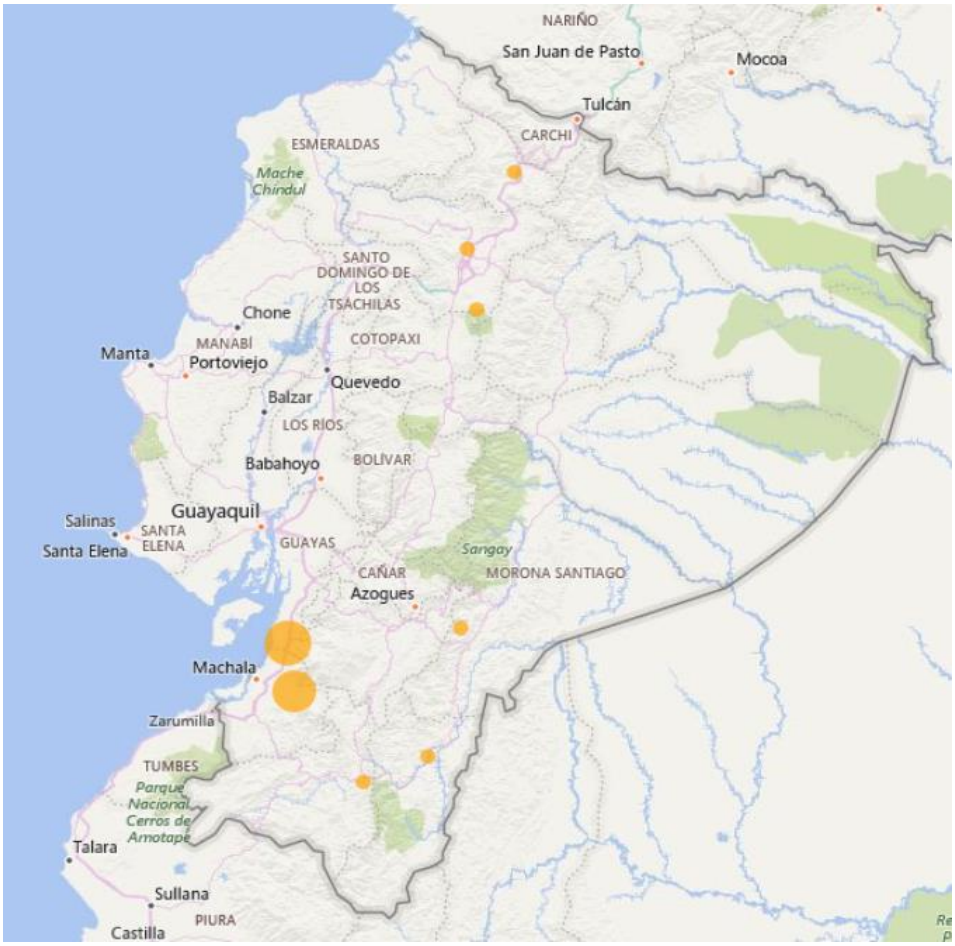
Fuente: El autor, histórico de precios investing.com, 2019.

Tabla 1.1.1.1 Producción de oro en el Ecuador por provincias

Producción de Oro por provincias según estadística de producción, Ministerio de Minería año 2016		
Provincia	Producción [Onzas Troy]	Producción [gramos]
Azuay	101389.3	3153559.72
Cotopaxi	588.9	18317.35
El Oro	97503.1	3032685.63
Imbabura	5358.6	166670.37
Loja	3230.0	100464.95
Morona Santiago	252.1	7839.76
Pichincha	135.0	4200.00
Zamora Chinchipe	8929.3	277732.80

Fuente: Ministerio de Minería del Ecuador, 2018

Figura 1.1.1.3. Mapa de densidad de producción de Oro en el Ecuador



Fuente: Ministerio de Minería, 2018

Los datos de la tabla 1.1.1.1 es posible visualizarlos en un mapa de densidades figura 1.1.1.3 donde el tamaño de los círculos está en proporción directa a la mayor o menor producción de oro en el Ecuador.

1.1.2 Trabajos Anteriores.

Se han realizado algunos trabajos de predicción del precio del oro y muchos otros trabajos relacionados a la predicción del precio de instrumentos financieros de inversión, ya sea, por medio de series de tiempo con modelos ARIMA, donde cada trabajo menciona diferentes parámetros de diferenciación, auto correlación y auto correlación parcial y muchos de los trabajos más actuales están usando redes neuronales artificiales.

En el pronóstico de precio del oro se pueden usar modelos de análisis de series de tiempo, como el trabajo de Banhi Guha y Gautam Bandyopadhyay " Gold Price Forecasting using ARIMA model (2016), donde determinan cada uno de

las funciones de auto correlación y auto correlación parcial en nueve diferentes modelos con p=1 , d= 0, 1 y q=1,2,3, tal que, d es el número de diferenciaciones realizadas al conjunto de datos para que estos sean estacionarios, el factor p es el orden de proceso autorregresivo y el factor q es el orden de procesos de medias móviles. Al comparar estos modelos considerando el criterio de información Bayesiano (BIC), los autores determinaron que la mejor combinación era un ARIMA (1,1,1).

En el Journal de Economía, Negocios y Gerenciamiento: Hossein Mombeini y Abdolreza Yazdani en “Modeling Gold Price vía Artificial Neural Network” (2015) realizaron un modelo de predicción del precio del Oro queriendo contrastar y verificar la efectividad de los modelos de análisis de serie de tiempo ARIMA a través de tres medidas de rendimiento como son R-squared, media cuadrática del error (RMSE) y el error absoluto medio (MAE). En ese trabajo, el modelo mejor ajustado fue un ARIMA (1,1,0) y, además, se hace una introducción a un modelo de redes neuronales (ANN), el cual consistió en una estructura de tres capas: input, capas escondidas y el output donde la capa media no tiene límites teóricos en el número de capas escondidas, pero generalmente habrá una o dos. Además, usaron un MLP (multi-layer perceptron network), que goza de una gran capacidad en términos de predicciones económicas y de variables financieras. Para este trabajo se recolectaron datos desde abril del 1990 hasta julio de 2008, lo que incluye 220 observaciones mensuales del precio del Oro por onza. Como una parte del desarrollo del modelo de predicción se tomaron como variables de entrada 7 parámetros como son: el índice del dólar, índice de inflación, precio del petróleo, índice de interés, el stock market y la producción mundial de oro. Ese estudio comprendió de 220 vectores de entrada y sus correspondientes vectores de salida de la data histórica, la cual se dividió (ANN) en dos partes: datos de entrenamiento y datos de validación, estos resultados mostraron que el método ANN tuvo mayor desempeño predictivo que el modelo ARIMA.

Adicional a los mencionados existen otros autores que han estudiado el uso de redes neuronales para predicción, así como de modelos ARIMA.

1.2. Descripción del proyecto.

Como hemos mencionado el precio del oro juega un rol importante en los sistemas económicos y monetarios, de hecho, el precio del oro y de otros activos están estrechamente relacionados. Por ejemplo, deberemos inferir en este trabajo que el precio del oro está relacionado con el valor del dólar, así como con el precio del petróleo.

Actualmente, el Ecuador está en un periodo de gran expectativa por los grandes proyectos mineros a desarrollarse, cuyo mineral principal de explotación es el oro. Estos proyectos deben necesariamente conocer al menos una tendencia, un pronóstico o acercamiento lo más preciso posible del precio del oro, para poder realizar la programación de sus operaciones e inversiones en base a los escenarios posibles.

Este pronóstico ayudará a prever flujos financieros, rentabilidades, programación de inversiones, planes de producción y de las operaciones, circunstancias y tendencias en el futuro de instrumentos financieros como las ETF´s, dando información útil a accionistas e inversores reduciendo el riesgo. Este trabajo definitivamente dará una herramienta de gerenciamiento y programación al inversor y al accionista. Por lo cual, será necesario buscar el mejor modelo de predicción usando diversas medidas de error.

1.2. Objetivo general.

Crear modelos estadísticos que permitan predecir el precio del oro de manera promedio mensual.

1.4. Objetivos específicos.

1.-Analizar mediante estadística descriptiva la evolución del precio del oro.
2.-Analizar las variables económicas que pueden influir en el precio del oro.
3.-Crear modelo de series temporales para pronosticar el precio del oro.
4.-Crear modelos relacionales para pronóstico del precio del oro de manera mensual.
5.- Introducir y modelar el precio del oro por medio de Redes Neuronales Artificiales en plataforma computacional.
6.- Comparar las metodologías utilizadas en el pronóstico del precio del oro.

1.5. Metodología.

Para este trabajo se usarán los pasos considerados en las metodologías de Ciencia de Datos (Data Mining and predictive analytics, 2015). Se debe empezar por adquirir los datos, identificar, recuperar y/o consultar cada uno de los índices propuestos en las fuentes financieras y económicas de los Estados Unidos. Con la data ya adquirida se procederá a realizar una exploración empezando por verificar su calidad y formato. Se deberá realizar un filtrado, ordenamiento y empaquetamiento de la información para que la misma pueda ser manejada por los softwares computacionales aplicados (Python). La estadística descriptiva o el sumario estadístico mostrará el comportamiento del grupo de datos, a la vez se realizará una visualización gráfica de la data. Debido a la gran cantidad de datos diarios generados desde 1970 se reordenará la data en base a promedios mensuales del precio del oro, esto debido a que para el manejo de análisis donde se use como referencia el precio del oro, es más propicio usar estos pronósticos para determinar las tendencias futuras y ejecutar las planificaciones respectivas; esto no quiere decir que el análisis no pueda ser considerado de manera semanal o diaria, pero requerirán de una mayor capacidad computacional y un mayor riesgo, así también, el incremento del horizonte de pronóstico debe ser lo más corto posible

Con los datos preparados, se usarán técnicas de análisis de predicción como regresiones, modelos relacionales, modelos de serie de tiempo ARIMA estacionales y no estacionales donde se encontrarán los parámetros de diferenciación, las funciones de auto correlación (ACF) y las funciones de auto correlación parciales (PACF) (Francisco Vera, apuntes de clases pronóstico de la demanda, 2017). Para la elaboración de los modelos ARIMA se tomará la data preprocesada para observar las estacionalidades y tendencias que la sucesión de datos están conformando, es necesario convertir la serie en una serie estacionaria, por lo que se decidirá las transformaciones a aplicar ya sea log y/o diferenciación, una vez que la serie es estacionaria se realizan los gráficos ACF y PACF para determinar la correlación de los datos a datos

pasados, con el fin de determinar los factores de autocorrelación e ir modelando nuestra serie.

Para el modelo relacional se escogieron factores macroeconómicos que puedan tener alguna influencia en el precio internacional del oro, tales como:

- Índice del dólar.
- Tasas de inflación.
- Precio del Petróleo
- Tasas de interés
- Stock Market index.
- Producción de oro mundial.

Para realizar el modelo relacional mediante Python, se utilizó una metodología llamada “Backward Elimination” o una eliminación en retroceso de las variables independientes, con el fin de obtener el modelo de regresión lineal multiple final (kirill Eremenko & Hadelin de Ponteves, Machine Learning on Python, Super Data Science, 2017).

El método usado para estimar el modelo de regresión lineal multiple será el método llamado OLS (ordinary Least Squares), este método busca minimizar la suma de las distancias al cuadrado entre los puntos observados y la línea de regresión.

Para realizar una eliminación de variables en retroceso se empieza eligiendo un nivel de significancia para mantenerse en el modelo, en nuestro caso NS=0.05, luego se realiza o se ajusta el modelo con todas las variables independientes (predictores) por analizar, con la tabla de resumen del modelo, se busca el predictor con el valor P más alto, si el valor P es mayor que el nivel de significancia (0.05) entonces eliminamos tal predictor del modelo, luego ajustamos el modelo con los predictores restantes, realizamos un bucle iterativo hasta que todos los valores P sean menores que 0.05

En la introducción de un modelo de redes neuronales (ANN) se codificará un algoritmo de generación de una red neuronal con célula de memoria para el manejo de los datos esta metodología se denomina LSTM (redes Long short term memory).

Las redes LSTM, son un tipo especial de redes neuronales, capaces de aprender dependencias a largo plazo. Fueron introducidos por Hochreiter y Schmidhuber (1997).

Los LSTM están diseñados explícitamente para evitar el problema de dependencia a largo plazo. Recordar información durante largos períodos de tiempo es prácticamente su comportamiento predeterminado. Esto las hace capaces y por ende muy usadas en los trabajos de series de tiempo. En este modelo se usará una red neuronal compuesta de dos capas de neuronas 300 y 10 y usando la misma base de datos conformada en el proceso de ordenamiento de datos.

Una vez construidos los modelos serán comparados a través de medidas de bondad de ajuste como la Suma cuadrática del Error o de validación cruzada.

.

CAPÍTULO 2

MARCO TEORICO

2.1. Factores e índices económicos que podrían tener alguna influencia en el precio del oro.

Se han considerado para nuestro ejercicio seis factores y/o índices económicos con los cuales se determinará si presentan influencia en el precio internacional del oro, estos son (Banhi Guha & Gautam Bandyopadhyay,2016):

- El índice del dólar.
- La tasa de inflación.
- Precio del Petróleo.
- Las tasas de interés.
- Índice del Stock Market.
- La producción mundial de oro.

Los factores tales como el índice del dólar, la tasa de inflación y las tasas de interés son los correspondientes a los Estados Unidos de América, debido a que el metal cotiza en su divisa, esta relación está en función del costo de oportunidad de mantener los lingotes que no generan rendimientos (Investopedia, 2019).

2.1.1. El índice del dólar**.**

El índice US dólar (USDX) es una medida del valor del dólar relativo al valor de una canasta de las monedas de los principales países que son socios comerciales de los Estados Unidos (James Chen, 2019).

El índice es calculado tomando como referencia las tasas de cambio de las 6 principales monedas mundiales que son el euro, el yen japonés, el dólar canadiense, la libra esterlina, la corona sueca y el franco suizo. El euro tiene la mayor ponderación frente al dólar, aproximadamente el 58% de la ponderación y el yen el 14%.

Si queremos interpretar el valor del índice, un valor de índice de 120% nos quiere decir que el dólar estadounidense se ha apreciado un 20% en relación con la canasta de monedas, es decir, si el índice del dólar aumenta, el dólar estadounidense está ganando fuerza o valor en comparación con las otras monedas. Así mismo si por ejemplo el índice es 80 habría una depreciación del 20% (Investopedia, 2019)

- EURO (EUR). El valor del Euro respecto al dólar, es decir, EUR/USD es el valor más importante de los que componen el índice con un peso del 57.6%.
- Yen Japonés (JPY). Es el valor de la divisa japonesa respecto al dólar, ocupa el segundo valor más importante en el índice que estamos analizando con un peso del 13.6%.
- Libra esterlina (GBP). Es el valor de la divisa inglesa, es la tercera que más peso tiene en el índice con un peso del 11.9%.
- Dólar Canadiense (CAD). Es la cuarta divisa que más peso tiene en el índice del dólar con un 9.1%.
- Corona Sueca (SEK). La corona sueca es la penúltima que más peso tiene en el USDX con un 4.2%.
- Franco Suizo (CHF). Esta divisa es la que menos relevancia tiene en el US dólar índex con un 3.6%.

2.1.2. Tasas de inflación.

La inflación (James Chen, 2019) es una métrica, comúnmente, porcentual de la tasa a la que se incrementa el nivel de precios promedio de una canasta de bienes y servicios determinados o seleccionados en una economía en un periodo de tiempo. La inflación puede indicar una disminución del poder adquisitivo de una moneda. Cuando los precios de estos bienes o servicios se incrementan, generalmente el banco central toma las medidas necesarias para mantener el nivel de inflación dentro de los límites permisibles y tener una economía nacional sin problemas.

Los efectos de la inflación en una economía son diversos y pueden ser tanto positivos como negativos. Los efectos negativos de la inflación incluyen la

disminución del valor real de la moneda a través del tiempo, la disminución de la intención de ahorro y de la inversión debido a la incertidumbre sobre el valor futuro del dinero y la escasez de bienes. Los efectos positivos incluyen la posibilidad de los bancos centrales de los estados de ajustar las tasas de interés nominal con el propósito de mitigar una recesión y de fomentar la inversión en proyectos de capital no monetarios (James Chen, 2019).

2.1.3. Precio del petróleo Brent Oíl.

El crudo Brent, (James Chen, 2019) es uno de los tres principales tipos de petróleo o crudo usado como punto de referencia para aquellos que realizan negociaciones de contratos de petróleo, futuros y derivados. Los otros dos tipos de petróleo de referencia son el WTI (West Texas intermediate) y el Dubái. El WTI es más dulce que le Brent y tiene un contenido de azufre más bajo.

Un barril de petróleo son 180 litros de petróleo (42 galones). Debido a las características actuales de la economía mundial el precio puede oscilar o variar con relación a los tiempos de prosperidad, niveles de consumo, la especulación, cantidad de reservas disponibles y acontecimientos sociales importantes, sobre todo en aquellos países productores y consumidores. A lo largo de la historia los precios han oscilado dentro de un rango de los 27 a 146 USD dólar por barril aproximadamente (Investopedia, 2019).

2.1.4. Tasas de interés.

La tasa de interés (Julia kagan, 2017), tipo de interés o precio del dinero, es la cantidad cobrada, como un porcentaje del capital, al prestatario por el uso de activos. También puede decirse que es el interés de una unidad de moneda en una unidad de tiempo o el rendimiento de la unidad de capital en la unidad de tiempo.

La crisis financiera de 2008 y la gran recesión han llevado algunas tasas de interés en muchas zonas del mundo a niveles cercanos a cero e incluso a intereses negativos (Julia Kagan, 2017).

2.1.5. Stock Market.

El Dow Jones (Will Kenton, 2019) es el promedio ponderado del precio de 30 acciones significativas negociadas en la Bolsa de Nueva York (NYSE) y el Nasdaq. El Dow Jones fue inventado por Charles Dow en 1896.

Conocido a menudo como "el Dow", es uno de los índices más antiguos y observados del mundo, e incluye compañías como Walt Disney Company, Exxon Mobil Corporation y Microsoft Corporation. Cuando las redes de televisión dicen que "el mercado está en alza hoy", generalmente se refieren al Dow.

El Dow es un índice de precios ponderados. Esto significa que las acciones con mayores precios de las acciones tienen mayor peso en el índice. Al inicio del Dow, Charles Dow calculó el promedio sumando los precios de las 12 acciones del componente Dow y dividiendo para 12. A lo largo del tiempo, ha habido adiciones y restas al índice, tales como fusiones y divisiones de acciones que debían ser contabilizadas en el índice. Cuando ocurre uno de estos eventos, el divisor para el Dow se ajusta para que el valor del índice no se vea afectado. Esta es la razón por la que el Dow puede alcanzar los 26,000, mientras que la suma total de los precios de las acciones de los componentes no se acerca a ese número (Investopedia, 2019).

2.1.6. producción mundial de oro.

La producción mundial de oro de acuerdo con el WGC, (World Gold Council, 2018) ha sido clasificado en tres áreas de provisión según su origen, producción minera netamente, Net producer Hedging y el oro reciclado (figura #7).

Figura 2.1.6.1 Oferta de Oro según la industria.

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Supply									
Mine production	2,748.5	2,857.5	2,929.0	3,110.5	3,202.6	3,290.0	3,398.7	3,446.7	3,500.9
Net producer hedging	-108.8	22.5	-45.3	-27.9	104.9	12.9	37.6	-25.5	-13.3
Recycled gold	1,679.1	1,651.1	1,670.8	1,247.7	1,187.8	1,121.4	1,281.5	1,156.1	1,167.5
Total supply	**4,318.8**	**4,531.1**	**4,554.5**	**4,330.3**	**4,495.4**	**4,424.3**	**4,717.8**	**4,577.3**	**4,655.1**

Fuente: Base de datos de WGC, 2019.

Es posible observar una tendencia creciente de la producción minera de oro mediante producción minera.

2.2. Análisis predictivo de datos.

El análisis predictivo de datos (UC San Diego, taller de ciencia de datos, 2018) es el arte de construir y usar modelos que hacen predicciones basado en patrones extraídos de datos históricos. Aplicaciones de análisis de datos predictivos incluyen:

Predicción de precios: La dinámica en tecnologías y la globalización de los negocios obligan a las empresas a ajustar constantemente sus precios para maximizar sus beneficios comparando y relacionando factores como cambios estacionales, cambios de demanda, etc.

Los modelos de análisis predictivo pueden ser entrenados para predecir precios óptimos basados en registros de ventas Las empresas pueden utilizar estas predicciones como una entrada en sus precios decisiones estratégicas.

Predicción de dosis: los médicos y científicos con frecuencia deciden qué cantidad de medicamento u otro químico para incluir en un tratamiento. Los modelos de análisis predictivo se pueden usar para ayudar a esta toma de decisiones al predecir las dosis óptimas basadas en datos sobre el pasado, dosis y resultados asociados.

Evaluación de riesgos: el riesgo es uno de los principales factores de influencia en casi todas las decisiones que una organización hace. Los modelos de análisis predictivo pueden usarse para predecir el riesgo asociado con decisiones tales como emitir un préstamo o suscribir una póliza de seguro. Estos modelos están entrenados usando datos históricos de los cuales extraen la clave de los indicadores de riesgo. El resultado de los modelos de predicción de riesgos puede ser utilizado por las organizaciones para hacer mejores juicios de riesgo.

Predicción de probabilidades: la mayoría de las decisiones empresariales se facilitarían mucho más si pudiéramos predecir la probabilidad o propensión de los clientes individuales a tomar diferentes acciones. El análisis predictivo de datos se puede usar para construir modelos que predicen acciones futuras de los clientes basados en el comportamiento histórico. Aplicaciones exitosas del modelado de propensión incluye predecir la probabilidad de que los clientes abandonen un operador de telefonía móvil a otro operador diferente, para

responder a los esfuerzos particulares de marketing, o para comprar diferentes productos.

Diagnóstico: En cada una de las diferentes áreas de la ciencia, los diversos profesionales como, médicos, ingenieros, genetistas, biólogos, etc., regularmente hacen diagnósticos los cuales se basan en su capacitación, habilidad y experiencia. Los modelos de análisis predictivos pueden ayudar a los profesionales a realizar mejores diagnósticos aprovechando una gran cantidad de ejemplos (data) histórico, más ejemplos o casos de los que un profesional podría genera en toda su carrera profesional. Los diagnósticos hechos por los modelos analíticos predictivos generalmente se convierten en una entrada al diagnóstico existente del proceso profesional (UC San Diego, taller de ciencia de datos, 2018).

2.3. Pronósticos.

En una estimación, (Larose, T. Daniel, 2015) se busca aproximar el valor de una variable numérica objetivo, usando un set de variables predictores sean estas numéricas y/o categóricas. En un pronóstico radica el proceso de estimación, fundamentado en las ciencias estadísticas. Ejemplos en los negocios y la investigación incluyen el pronóstico del precio de una acción meses en el futuro, porcentaje de incremento en las muertes en accidentes de tránsito si el límite de velocidad es superado.

Lograr una predicción, sea cual sea el fin, con precisión no es un proceso fácil, ya que pueden existir muchos factores internos y externos relacionados al valor por pronosticar, políticos, demográficos, económicos, que afectan a la predicción. Por esta razón es necesario contar con un desarrollo de la ciencia de datos, al mismo ritmo de desarrollo de la tecnología computacional actual.

Al comparar el mundo real con las predicciones realizadas aun existirán ciertas variaciones, para esto es necesario en paralelo ir desarrollando metodologías operacionales para absorber estas variaciones, como podría ser tener capacidad adicional, el manejo de inventarios de seguridad, la posibilidad de reprogramar pedidos, pero hay que considerar que las grandes variaciones sí podrían afectar de gran manera las decisiones operacionales de una empresa.

Los métodos de predicción se han desarrollado mediante metodologías clásicas de las cuales existen dos tipos para pronóstico: series de tiempo y causales o de clasificación. La lógica básica de todos los métodos de predicción es que los datos del pasado y los patrones de datos son indicadores confiables para predecir el futuro. En estos casos, los datos del pasado se procesan mediante las metodologías estadísticas aplicadas a las series de tiempo para hacer un pronóstico (Larose, T. Chantral, 2015).

2.3.1. Pronósticos por series de tiempo.

En las diferentes ramas científicas y sociales como la ingeniería, economía, medicina, política, migraciones, etc., los datos son generados y tomados siguiendo una temporalidad equitativa, por ejemplo, hora, días, mes, año. Los modelos de series de tiempo pueden ser del tipo univariantes o multivariantes (Peña, 2001).

Las series de univariantes es donde se analiza una sola variable temporal en base a su propio pasado; las series multivariantes se analizan varias series temporales a la vez, cuando se construyen este tipo de modelos, en muchos casos se supone cierta relación entre los datos históricos de cada una de las variables.

En el momento de plantear un análisis de series de tiempo se lograrán o se tendrán dos objetivos muy claros como, describir las características estadísticas de la serie tales como tendencias, comportamiento estacional, estos dados en consistencia con algunos datos estadísticos como son media, varianza, etc. Otro de los objetivos es de realizar la predicción de los valores futuros de las variables.

Los valores de una serie de tiempo pueden dividirse en factores o términos tales como nivel promedio, tendencia, estacionalidad, ciclo y error. Cuando se suman los componentes (o en algunos casos se multiplican), serán iguales a la serie de tiempo original.

La estrategia básica que se utiliza en los pronósticos por series de tiempo es identificar la magnitud y la forma de cada uno de los componentes basándose en los datos disponibles. Estos componentes (con excepción del componente aleatorio), se proyectan hacia el futuro. Si sólo queda un componente aleatorio

pequeño y el patrón persiste en el futuro, se obtendrá un pronóstico confiable (Jiménez Daniela, 2011).

La descomposición de una serie de tiempo es la siguiente:

$$y(t) = (a + b_t)[f(t)] + e$$

(2.1)

donde:

y (*t)*: demanda durante el período t

a: nivel

b: tendencia

f (*t*): factor de estacionalidad (multiplicativo)

e : error aleatorio

En la ecuación generalizada de serie de tiempo se tiene un nivel, tendencia, estacionalidad y error aleatorio. Cada uno de estos términos se estima a partir de datos del pasado para desarrollar una ecuación que se utiliza entonces para pronosticar la demanda a futuro (Jiménez Daniela, 2011).

En pronósticos por series de tiempo, se usan los siguientes símbolos y terminologías:

D_t : Demanda durante el periodo t
F_{t+1} : Demanda pronosticada para el periodo $t + 1$
$e_t = D_t - F_t$: error de pronóstico en el periodo t
A_t : promedio calculado hasta el periodo t

2.3.2. Técnicas más usadas en pronósticos de series de tiempo.

2.3.2.1 Promedio móvil.

Esta es sin duda una de la metodología más simple al realizar pronósticos por series de tiempo (José Ángel Fernández, 2003). En este método se supone que la serie de tiempo tiene sólo un componente de nivel y un componente aleatorio. No se asumen patrones de estacionalidad, tendencias ni componentes de ciclos en datos de la demanda.

Cuando se utiliza el promedio móvil se selecciona un número dado de periodos N para los cálculos. Después se calcula el promedio P_t para los N periodos del pasado al momento t de la manera siguiente

$$P_t = \frac{D_t + D_{t+1} + \cdots + D_{t-N+1}}{N}$$

(2.2)

Como se supone que la serie de tiempo es horizontal, el mejor pronóstico para el periodo *t + 1* es simplemente una continuación de la demanda promedio observada a lo largo del periodo t. De esta manera se obtiene:

$$F_{t+1} = P_t$$

(2.3)

Cada vez que se calcula F_{t+1}, el valor más reciente se incluye en el promedio y se quita la observación de la demanda más antigua. Este procedimiento mantiene un número N de periodos dentro del pronóstico y permite que el promedio se mueva conforme se observan los nuevos datos.

Como regla general, mientras más largo sea el período en que se hace el promedio, más lenta será la respuesta ante los cambios en la demanda. Los periodos más largos tienen, por lo tanto, la ventaja de dar estabilidad a los pronósticos. Sin embargo, también tiene la desventaja de responder con mayor lentitud a los cambios verdaderos en el nivel de demanda, El analista y la velocidad de respuesta al seleccionar una compensación apropiada entre la estabilidad y la velocidad de respuesta al seleccionar la longitud de N que la longitud de N que se utilizará para el promedio (Jiménez Daniela, 2011).

Una manera de hacer que el promedio móvil responda con mayor rapidez a los cambios de una demanda, por ejemplo, es colocar un peso relativo superior sobre la demanda reciente en vez de hacerlo sobre la demanda más antigua. Esto se denomina promedio móvil ponderado y se calcula como sigue:

$$F_{t+1} = P_t = W_1 D_t + W_2 D_{t-1} + \cdots + W_N D_{t-N+1}$$

(2.4)

Con la condición de que

$$\sum_{i=1}^{N} W_i = 1$$

(2.5)

Con el promedio móvil ponderado se puede especificar cualquier peso deseado siempre y cuando su suma sea igual a 1. Además, el promedio móvil simple es solamente un caso especial de promedio móvil ponderado en el que todos los pesos son iguales:

$$w_i = \frac{1}{N}$$

(2.6)

Unas de las desventajas del promedio móvil ponderado es que debe utilizarse toda la historia de la demanda de los periodos N junto con el cálculo. Además, la respuesta de un promedio móvil ponderado no puede cambiarse con facilidad sin alterar cada uno de los pesos específicos. Para resolver estas dificultades, se ha desarrollado el método de suavización exponencial (Jiménez Daniela, 2011).

2.3.2.2. Modelos ARIMA y suavización exponencial.

La suavización exponencial (Jiménez Daniela, 2011) se basa en la idea, muy simple, de que es posible calcular un promedio nuevo a partir de un promedio anterior y también de la demanda más recientemente observada.

Para formalizar el razonamiento anterior se escribe:

$$\boldsymbol{A_t} = \propto \boldsymbol{D_t} + (\mathbf{1} - \propto)\boldsymbol{A_{t-1}}$$

(2.7)

Donde $\boldsymbol{A_{t-1}}$ es el promedio anterior, $\boldsymbol{D_t}$ es la demanda que se acaba de observar y $\propto$ es la proporción del peso que se da a la demanda nueva contra la que se le da al promedio anterior (*0 <= α <= 1*).

Si se desea que responda en alto grado a la demanda reciente, se debe elegir un mayor valor para α. Si se desea que $\boldsymbol{A_t}$ responda con mayor lentitud, entonces α será más pequeña. En la mayor parte del trabajo de pronósticos α recibe un valor que se encuentra entre 0.1 y 0.3 para que conserve una estabilidad razonable.

En la suavización exponencial simple, como en el caso de los promedios móviles, se supone que la serie de tiempo es plana, que no tiene ciclos y que no existen componentes de estacionalidad ni tendencia. Entonces, los pronósticos de suavización exponencial para el siguiente periodo serán simplemente el promedio obtenido hasta el período actual. Es decir,

$$F_{t+1} = A_t$$

(2.8)

En este caso el pronóstico también elimina un periodo del promedio suavizado. Se puede sustituir la relación anterior en la ecuación (7) para obtener la siguiente ecuación:

$$\boldsymbol{F_{t+1}} = \propto \boldsymbol{D_t} + (\mathbf{1}-\propto)\boldsymbol{F_t}$$

(2.9)

En ocasiones esta forma alterna de suavización exponencial simple (o de primer orden) es más fácil de usar que la ecuación (7) debido a que utiliza pronósticos en lugar de promedios (Jiménez Daniela, 2011).

Otra manera de considerar la igualación exponencial es reacomodar los términos del lado derecho de la ecuación (8) para obtener:

$$\boldsymbol{F_{t+1}} = \boldsymbol{F_t} + \propto (\boldsymbol{D_t} - \boldsymbol{F_t})$$

(2.10)

En esta forma se indica que el pronóstico nuevo sería el pronóstico anterior más una proporción del error entre la demanda observada y el pronóstico anterior. Se puede controlar la proporción de error utilizada mediante la elección de α.

El nombre “suavización exponencial” se puede explicar si se escribe la ecuación (8) en términos de todas las demandas anteriores. Al sustituir Ft en la ecuación (8) se obtiene:

$$\boldsymbol{F_{t+1}} = \propto \boldsymbol{D_t} + (\mathbf{1}-\propto)[\propto \boldsymbol{D_{t-1}} + (\mathbf{1}-\propto)\boldsymbol{F_{t-1}}]$$

(2.11)

Después, la sustitución de $\boldsymbol{F_{t-1}}$en la ecuación anterior da:

$$\boldsymbol{F_{t+1}} = \propto \boldsymbol{D_t} + (\mathbf{1}-\propto) \propto \boldsymbol{D_{t-1}} + (\mathbf{1}-\propto)(\mathbf{1}-\propto)[\propto \boldsymbol{D_{t-2}} + (\mathbf{1}-\propto)\boldsymbol{F_{t-2}}]$$

(2.12)

Si se continua con esta situación se llegara a la expresión:

$$\boldsymbol{F_{t+1}} = \propto \boldsymbol{D_t} + (\mathbf{1}-\propto) \propto \boldsymbol{D_{t-1}} + (\mathbf{1}-\propto)^{\mathbf{2}} \propto \boldsymbol{D_{t-2}} + \cdots + (\mathbf{1}-\propto)^{\boldsymbol{t-1}} \propto \boldsymbol{D_1} + (\mathbf{1}-\propto)^{\boldsymbol{t}} \boldsymbol{F_1}$$

(2.13)

Esta expresión indica que los pesos de cada demanda precedente disminuyen exponencialmente en un factor de (1- α), hasta que se alcance la demanda del primer periodo y del pronóstico inicial F1.

Nótese que el peso de la demanda disminuye exponencialmente con el tiempo y que todos lo pesos suman 1. Por lo tanto, la suavización exponencial es simplemente una forma especial que adquiere el promedio ponderado en donde el peso disminuye exponencialmente con el tiempo.
El procedimiento para seleccionar el valor de α resulta claro ahora. Se debe calcular el pronóstico para varios valores de α, si uno de los valores de α da como resultado un pronóstico que tenga un menor grado de desviación que los otros, entonces se prefiere este valor de α. Si no existe una elección obvia entonces debe hacerse un intercambio entre la tendencia y la desviación absoluta para elegir el valor de α que se prefiere.

Desafortunadamente, la suavización exponencial no siempre puede utilizarse en la práctica debido a la tendencia que tienen los datos a mostrar variaciones de acuerdo con las estaciones. Cuando se presentan estos efectos pueden utilizarse suavizaciones de segundo orden, de tercer orden, de tendencia corregida o de estacionalidad (Jiménez Daniela, 2011).

En los modelos de series de tiempo ARIMA (Peña, 2001) con descomposición mediante términos $Y_t = T_t + S_t + \varepsilon_t$, donde T es una componente de tendencia, la S es de estacionalidad y ε el error. Como parte importante es necesario poder estimar ε_t y determinar si es o no ruido blanco, es decir, que los errores no guarden una correlación estadística en dos tiempos distintos. Si se ha determinado que el ε_t no es ruido blanco, tocara modelar mediante tres posibles opciones. Medias móviles de orden q, MA(q), los modelos autorregresivos AR de orden p, AR(p) y las medias móviles autorregresivos ARMA(p,q).

Para empezar a trabajar estas series de tiempo, aplicando modelos autorregresivos y/o medias móviles, deberemos primero verificar la estacionalidad de la serie para esto es posible aplicar transformaciones a la serie siendo las transformaciones logarítmicas las más frecuentes para lograr la estacionariedad en varianza y las diferenciaciones para lograr la estacionariedad en media (Peña, 2001).

Una vez realizado esto es necesario realizar los gráficos de auto correlación ACF y auto correlación parcial PACF, estos son os que definirán el orden de los modelos AR(p), MA(q) o ARIMA(p,d,q) respectivo.

Generalmente (Daniel Pena, 2001) en los modelos ARIMA de series económicas los valores de d (número de diferenciaciones realizadas a la serie), más comunes suelen ser d=0,1,2 pero para poder determinar cuál de estos valores es el más adecuado a nuestro modelo en particular es necesario usar los siguientes instrumentos:

a) Realizar gráficos de la serie original y transformaciones que sean necesarias y verificar en qué punto la serie con las condiciones de estacionariedad.
b) Realizar los correlogramas de la serie original y las transformaciones para verificar si hay decrecimiento rápidamente hacia cero.
c) Realizar prueba de hipótesis de raíces unitarias (test Dickey -Fuller).

El literal c de hipótesis de raíces unitarias nos da un contraste estadístico, que nos permite realizar una inferencia de si hay o no una raíz unitaria en la serie, es decir, sobre la no estacionariedad de las series. Si se rechaza la hipótesis nula de existencia de raíz unitaria, no se realizará otra diferenciación, caso contrario si no se rechaza la hipótesis nula se hará una diferenciación más de orden 1.

De manera general si una serie Y_t es integrada de orden d, se pude representar matemáticamente mediante el siguiente modelo:

$$\varphi_p(L)\Delta^d Y_t = \delta + \theta_q(L)\Delta^d a_t$$

(2.14)

Donde:

$\varphi_p(L)$: polinomio autorregresivo estacionario.

$\theta_q(L)$: polinomio invertible de medias móviles.

p: orden del polinomio autorregresivo estacionario.

q: orden del polinomio de medias móviles invertible.

d: orden de integración de la serie o número de diferencias realizadas a la serie para que esta sea estacionaria.

Y_t: Observaciones en tiempo t.

a_t: Termino de error en tiempo t.

2.3.3. Técnicas para determinar los mejores pronósticos.

Una vez que se generan los modelos estadísticos para las series de tiempo es necesario determinar o tener una medida del rendimiento del modelo estadístico. Una de las medidas es simplemente la suma aritmética de todos los errores, con lo que se refleja la tendencia del método de pronóstico.

Otro ejemplo de medida del error de pronóstico es la desviación absoluta. En este caso se suma el valor absoluto de los errores, de tal manera que los errores negativos no cancelen a los positivos. El resultado es una medida de variación en el método de pronóstico. Si un pronóstico tiene tanto una tendencia como una desviación absoluta, resulta claro que se le debe preferir (Jiménez Daniela, 2011).

2.3.3.1. Técnicas de errores de pronósticos.

A continuación, se detallan las técnicas de pronóstico de error (Peña, 2001):
MSE (error medio al cuadrado), es una medida de la calidad de un estimador, este es un valor positivo o no negativo y siempre los valores más pequeños o alrededor de cero son los de mejor rendimiento. El MSE es la media de los errores al cuadrado.

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(Y_i - \overline{Y}_i)^2$$

(2.15)

El RMSE, es la raíz del error medio al cuadrado y una de las medidas que se usa con mayor frecuencia

$$RMSE = \sqrt{MSE}$$

(2.16)

Donde $(Y_i - \overline{Y_i})$ es el valor del error en un período t.

2.3.3.2 Pronósticos avanzados por series de tiempo.

Se puede ajustar cualquier modelo matemático deseado con una serie de tiempo como la que se muestra en la ecuación 2.14, con componentes de nivel, tendencia y estacionalidad. Por ejemplo, se puede ajustar un modelo mediante los métodos de regresión lineal o mediante el uso de métodos no lineales. En algunos casos, el modelo resultante puede brindar un pronóstico más exacto que la suavización exponencial. Sin embargo, resulta más costoso un modelo adaptado a las necesidades del usuario, por lo que debe hacerse una compensación mediante la exactitud y el costo del modelo.

Se desarrolló tiempo atrás el método Box-Jenkins para el pronóstico por series de tiempo. Este método tiene una fase especial para la identificación del modelo y permite un análisis más preciso de los modelos propuestos de lo que es posible con los demás modelos. El método Box-Jenkins, sin embargo, requiere de aproximadamente 60 periodos de datos del pasado y su uso resulta demasiado costoso para los pronósticos rutinarios de muchos artículos. Para un pronóstico especial de ventas en que se involucre una decisión costosa, sin embargo, quizás sea recomendable utilizar el Box-Jenkins.

En resumen, los métodos por series de tiempo son útiles para los pronósticos a corto o mediano plazo cuando se espera que el patrón de demandas permanezca estable. Los pronósticos por series de tiempo son con frecuencia insumos para decisiones que se relacionan con la planeación de producción agregada, presupuestos, asignación de recursos, inventarios y programación (Daniel Peña, 2001).

2.4. Machine Learning y redes neuronales artificiales.

Machine learning comprende un conjunto de algoritmos que tienen la capacidad de aprender de un conjunto de datos (Ian Goodfellow, Yoshua Bengio, & Aaron Courville, 2016). El termino aprender usado para un programa computacional que aprende de una cierta experiencia E con respecto a una tarea asignada T y

esto con una medida de rendimiento P, es decir, su rendimiento en la tarea T, medida por P, mejora con la experiencia E.

Entre las tareas más comunes que el machine learning puede realizar están:

- Clasificación.
- Clasificación con entradas faltantes.
- Regresiones.
- Transcripciones.
- Traducciones.
- Detección de anomalías.

Para el caso de las regresiones se realiza un análisis de datos y uso de un modelo para hacer una predicción como soporte técnico al tomar una decisión. En el análisis predictivo de datos, utilizamos una definición amplia de la palabra predicción. El uso del término predicción tiene en sí mismo un aspecto temporal: anticipar de cierta forma el futuro, la asignación de un valor a cualquier variable desconocida. Esto es, por ejemplo, predicción de precios o demandas de cierto tipo de productos y bajo el cual podrían partir otros tipos de análisis como optimización de transporte, inventarios, bodegaje, etc. Por lo tanto, un modelo es entrenado para hacer predicciones basadas en un conjunto de ejemplos históricos.

2.4.1. clasificación del machine learning.

Este conjunto de algoritmos que hemos mencionado puede ser ampliamente clasificados como supervisados y no supervisados (Brownlee, 2018).

El conjunto de algoritmos llamados supervisados, son cuando estos experimentan una base de datos que contiene un grupo de características, pero cada ejemplo, al que corresponden esas características, es también asociado a una designación u objetivo.

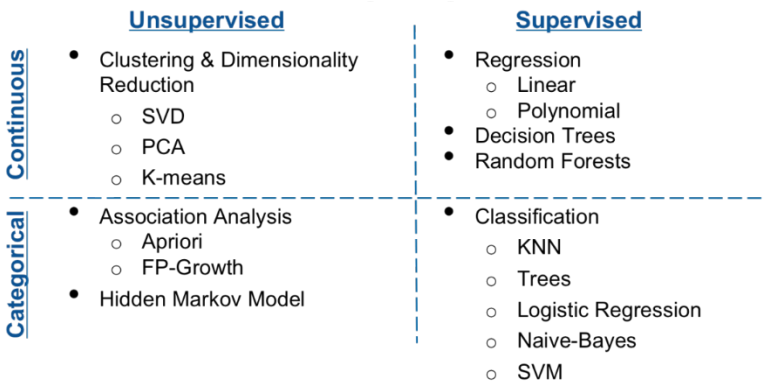
A su vez el conjunto de algoritmos no supervisados, cuando estos experimentan una base de datos que contienen muchas características, no se dan las designaciones u objetivos de manera total o parcial al algoritmo de aprendizaje, dejándolo por si solo para encontrar estructura en su input.

Machine Learning es un proceso automatizado que extrae patrones de los datos, en la figura 8 y 9, se hace una descripción generalizada del conjunto de algoritmos que están dentro del machine learning (Brownlee, 2018).

Figura 2.4.1.1 Algoritmos aplicados a Machine learning.

Fuente: Jason Brownlee, 2019.

Figura 2.4.1.2 Ejemplo de algoritmos de Machine Learning supervisados y no supervisados.

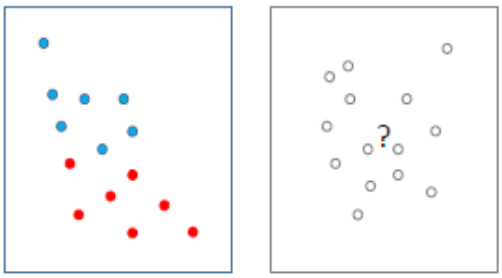


Fuente: Google, colab reserch, 2019

2.4.2 Redes neuronales artificiales.

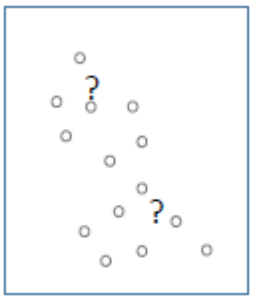
Dentro del concepto de machine learning existe una metodología llamada Deep learning, cuya estructura se asemeja al funcionamiento de las neuronas en el cerebro humano (Ian Goodfellow , Yoshua Bengio, & Aaron Courville, 2016). Deep learning está basado en la filosofía del conexionismo, es decir, mientras una sola neurona biológica o una sola característica en un modelo de machine learning no es inteligente, una gran cantidad de estas neuronas o características actuando juntas pueden mostrar un comportamiento inteligente. Uno de los factores clave responsable por la mejora que tienen en la exactitud las redes neuronales y la mejora en la complejidad de las tareas que pueden resolver, en el dramático incremento en el tamaño de las redes que usamos.

Figura 2.4.2.1 Ejemplificación grafica de aprendizaje de maquina supervisado y no supervisado



Fuente: Microsoft Deep Learning, 2019.

Figura 2.4.2.2 Ejemplificación de aprendizaje no supervisado



Fuente: Microsoft Deep Learning,2019.

Para nuestro caso de estudio, series de tiempo, tenemos el conjunto de algoritmos que están dentro del campo de aprendizaje supervisado.

Revisemos un poco la terminología aplicada en el machine learning en base al campo de interés. En el procesamiento de los datos se tendrá la tarea de realizar una regresión, donde básicamente este análisis empieza con la ecuación de la recta (Sayan Pathak, 2017).

$$y^* = mx + b$$

(2.17)

Input,

$$X = datos\ (feature)$$

(2.18)

Outputs,

$$y = etiquetas\ observados\ correspondientes\ a\ los\ datos\ (labels)$$

(2.19)

$$y^* = etiquetas\ predichas$$

(2.20)

Parámetros del modelo,

$$m : pendiente$$

$$b: intercepto$$

El proceso de este modelo es el aprendizaje de los parámetros m y b para lo que tendremos un entrenamiento con los datos, una validación y una prueba.

Para la etapa de entrenamiento se tiene una muestra de los datos $\{x, y\}_n$ (training set), donde sería posible por medio de un ajuste de los mínimos cuadrados tener una estimación de los parámetros m y b. Sin embargo, para nuestros fines necesitamos poder entrenar modelos de Deep learning. Una vez que tenemos nuestra muestra de datos y se toman las lecturas del set de observaciones *x* y sus correspondientes predicciones *y* (training set), ahora se tomarán los valores de x como valores de entrada o input y aplicaremos el modelo mx+b, donde los parámetros m y b serán asignados aleatoriamente m1 y b1, entonces tenemos *y*, *y**, una vez estos datos iterados procedemos a calcular el error o loss.

$$loss = \sum_{l=1}^{n} (y - y^*)^2$$

(2.21)

entonces,

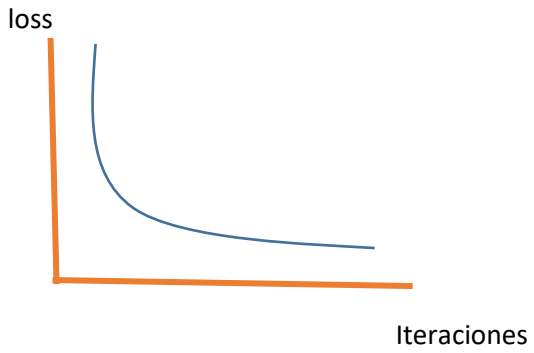
$$m1, b1 \rightarrow l1\ (loss\ 1)$$

$$m2, b2 \rightarrow l2\ (loss\ 2)$$

$$mn, bn\ \rightarrow l\ n\ (loss\ n)$$

Estos datos permiten obtener una relación entre el número de iteraciones y el loss o error donde deberemos intuitivamente seleccionar el menor error.

Figura 2.4.2.3 Medida de error vs iteraciones realizadas

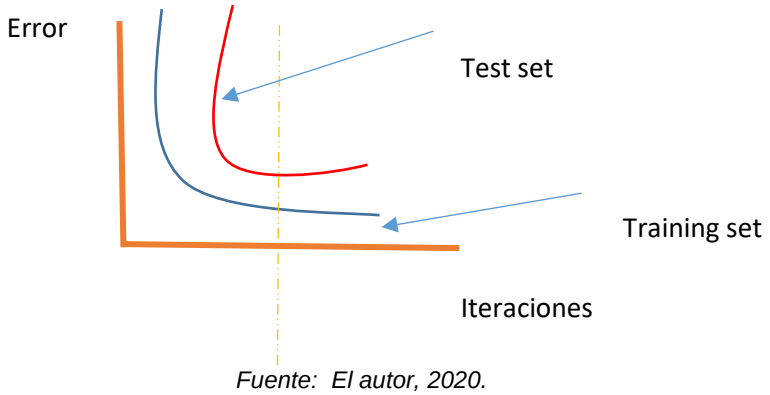


Fuente: El autor, 2020.

Al tener un valor bajo de error en los parámetros m_n, b_n, tendremos un modelo Z , el cual , trabaja y se ajusta muy bien a la data tomada como muestra para el training set , pero cuando los datos están fuera de esa muestra el error puede ser muy alto al usar el mismo modelo Z y lo deseable es que los datos fuera de la muestra de entrenamiento tengan un performance comparable al obtenido a la fase de entrenamiento.

Para la validación y test del modelo se procede a realizar una muestra de datos diferente a la de entrenamiento y se procede a realizar el reporte de error, donde debemos encontrar los parámetros que nos den los errores más bajos para ambas muestras (Sayan Pathak, 2017).

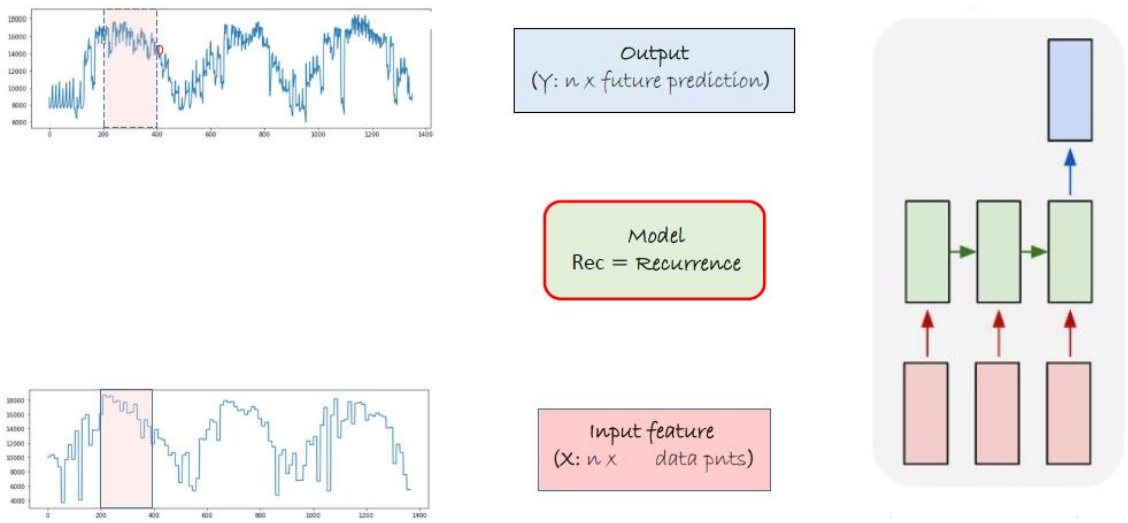
Figura 2.4.2.4 Criterio de selección de modelo acorde a el error vs # iteraciones



Fuente: El autor, 2020.

Para el desarrollo de redes neuronales que trabajen con series de tiempo es necesario podamos introducir el concepto de secuencias. Hay diferentes tipos de secuencias que son posibles de ejecutar. Para nuestro interés de manejar series de tiempo solo mencionaremos un par de ellas. La primera de ellas donde tenemos múltiples inputs entrando al modelo y pronosticaremos un valor futuro. En este caso según la figura 2.4.2.5 los bloques rosados son los inputs, los bloques azules son los outputs y los bloques verdes son los modelos que construiremos con recurrencia y es llamado como “many to one”.

Figura 2.4.2.5 Ejemplificación de secuencias en series de tiempo “Many to one”

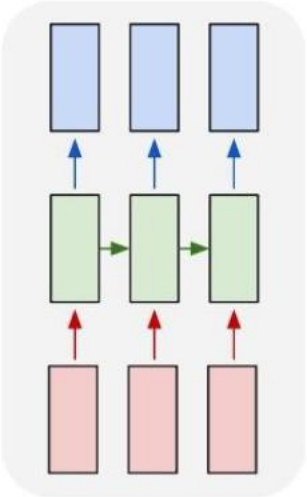


Fuente: Karpathy, Stanford, 2018.

Aplicado nuestro ejemplo de precios del oro, nosotros tenemos los registros de los precios al final de cada mes, donde se tomará una ventana de datos para realizar una predicción, se construirá el modelo y este debe ser un modelo recurrente.

Existe otro tipo de secuencias llamado “many to many” y está representado mediante la figura 15.

Figura 2.4.2.6 Secuencias de tipo “Many to Many”

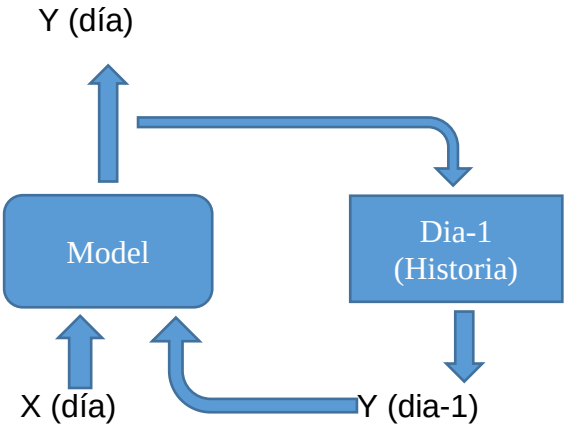


Fuente: Karpathy, Stanford, 2018.

Este tipo de estructura es utilizado cuando se requieren muchos outputs y cada uno tiene una correspondencia uno a uno, como es el caso de formación de oraciones donde cada una de las palabras son transformadas en números para su proceso informático. Este tema no será abordado.

En el problema propuesto de predicción de precio del oro podemos registrar un pronóstico mediante un modelo que se ajuste y nos dé un pronóstico en un punto en el tiempo, pero que pasaría si pudiésemos adicionar a ese modelo información de lo sucedido en periodos anteriores al momento del pronóstico, esta adición sin duda tendrá una gran influencia en el valor que sería en el momento predicho, esto es, el valor de la incorporación de la historia para hacer mejores pronósticos (Karpathy , 2015).

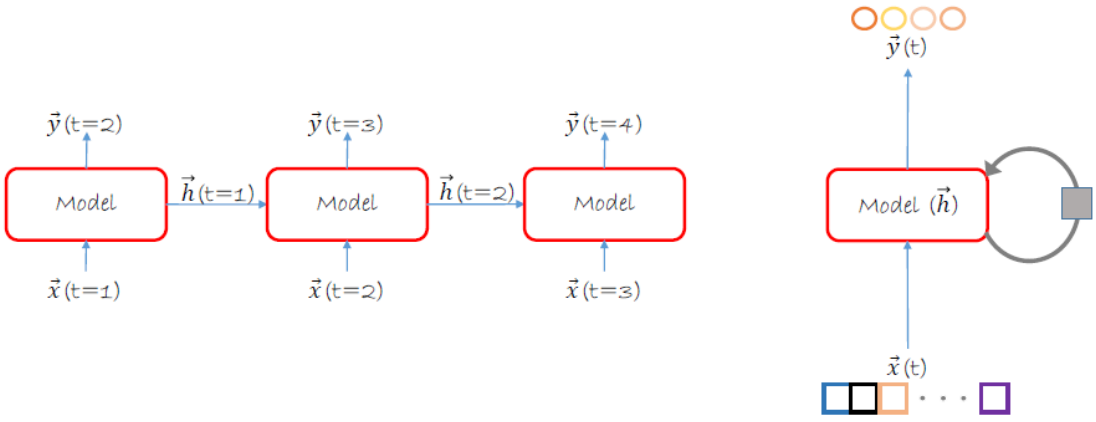
Figura 2.4.2.7 Incorporación de datos históricos al modelo



Fuente: El autor, 2019.

Como podemos incorporar historia en nuestro modelo de redes neuronales. Para esto en redes se tiene esto llamado el estado interno, el cual es un vector de dimensión m, cabe mencionar que el input x puede ser de igual manera un vector de dimensión n y el output y puede ser también un vector de dimensión c. Para cada vez que se realiza una predicción y, nosotros también admitiremos historia, es evidente que en $t = 1$, la historia de los datos será cero, pero subsecuentemente con el transcurrir del tiempo se tendrá una gran cantidad de datos históricos (Colab research, 2018). Entonces como la red recurrente se verá según figura.

Figura 2.4.2.8. Recurrencia.



Fuente: Microsoft Deep Learning, 2019.

$\vec{x}(t)$: input (vector n-dimensional) en tiempo t.

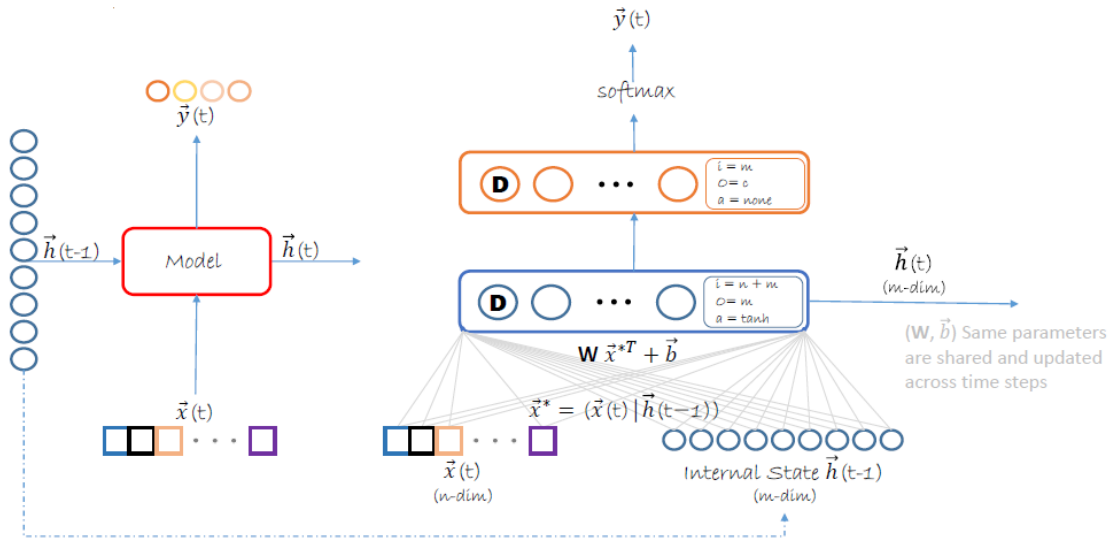
$\vec{y}(t)$: Output (vector c-dimensional) en tiempo t.

$\vec{h}(t)$: Estado interno (vector m-dimensional) en tiempo t

Acorde a la figura anterior se puede observar que tendremos un vector de input, un vector de output y un estado interno o historial que mantendrá actualizando cada etapa de recursión. Entonces esa recursión en cada etapa implica que hay una noción del tiempo y se verá en los modelos computacionales de ANN que se usaran términos tales como recursión y pasos de tiempo. Nótese las diferencias dimensionales de cada uno de los vectores que intervienen.

En la recurrencia generada en los modelos neuronales nuestro input en el tiempo t tendría una dimensión *(n+m),* dada tanto por la dimensión del vector de estado interno $\vec{h}(t-1)$ y el vector de datos $\vec{x}(t)$, este nuevo vector de entrada, $\overrightarrow{x^*} = (\overrightarrow{x}(t)|\vec{h}(t-1))$, debe realizar la operación básica, $W\,\overrightarrow{x^{*T}} + \vec{b}$, donde $(W, \vec{b})$, son parámetros que son compartidos y actualizados al paso del tiempo (Colab research, 2018).

Figura 2.4.2.9 Arquitectura de los modelos con Recurrencia.

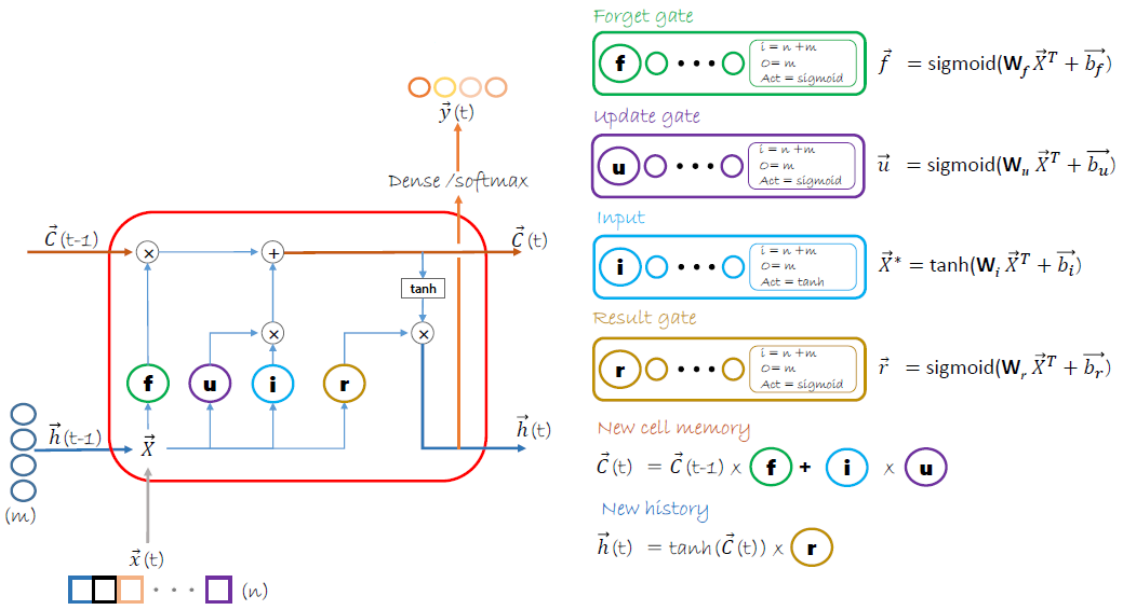


Fuente: Microsoft Deep Learning, 2019

Así que en el aprendizaje en una gran cantidad de datos y debido a que cada paso en el tiempo se comparte y se actualizan los parámetros $(W, \vec{b})$, es que

tenemos una limitación de memoria. Es aquí donde entra el concepto de celdas de memoria a corto y largo plazo “LSTM”.

Figura 2.4.2.10 Esquema de funcionamiento de células LSTM.



Fuente: Karpathy, Stanford, 2018.

Para la predicción de series de tiempo, los mejores algoritmos son los basados en Redes neuronales recurrentes y/o las denominadas LSTM o *“long short term memory”*. Las LSTM son un tipo particular de redes neuronales recurrentes que está siendo cada vez más trabajada dentro de la comunidad de machine learning. En forma simple, las LSTM tienen celdas de estado contextual que actúan como celdas de memoria de largo termino y corto termino.

El output de estas celdas esta modulado por el estado de esas celdas. Esto es una propiedad muy importante cuando se necesita la predicción de una red neuronal que depende del contexto histórico de los inputs, en lugar de solo depender del último de los inputs. A manera de ejemplo simple consideremos la siguiente serie de números 5,6, 7, ¿?, nos gustaría tener el siguiente valor de la serie que seria 8, *(x+1)*, sin embargo, si nos dan la siguiente serie de números 2,4,6, ?, el valor siguiente de la serie también seria 8, *(x+2).* El resultado de la predicción toma una connotación y rumbo diferente cuando tomamos en cuenta la información contextual de los datos anteriores y no solo la última (Karpathy, 2018).

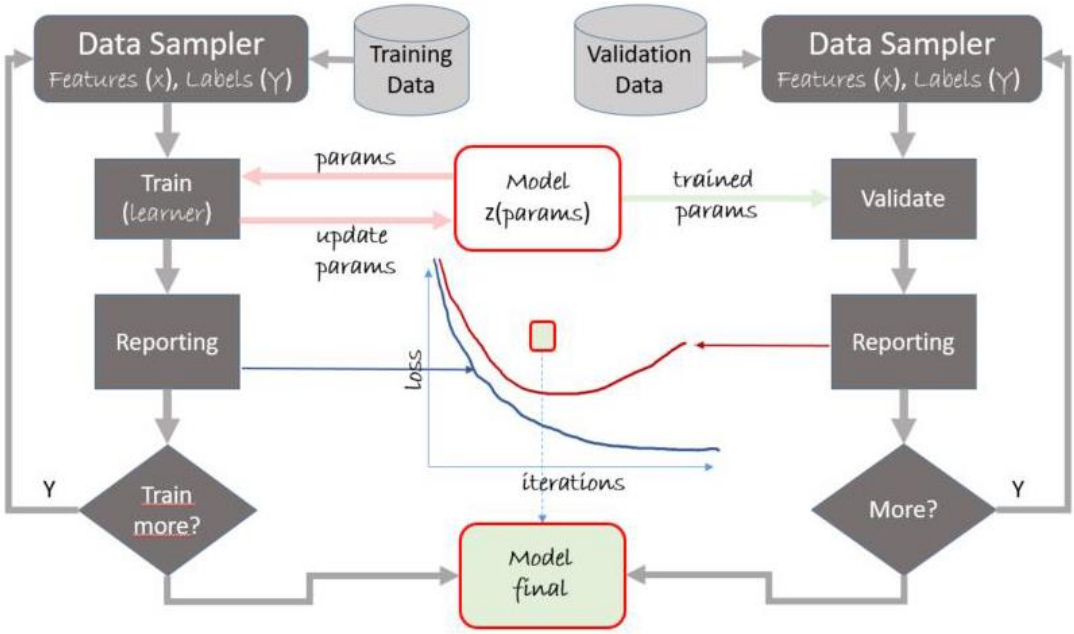
LSTM, logra mantener la información de los inputs integrando un bucle que permite que la información fluya de un paso al siguiente. Estos bucles hacen que las redes neuronales recurrentes parezcan mágicas.

Por otro lado, cuanto más tiempo haya pasado, es menos probable que el próximo output dependa de un input más antiguo. Esta vez, la distancia de dependencia es también una información contextual que debe aprenderse. Las redes LSTM lo gestionan aprendiendo cuando recordar y cuando olvidar (Colab research, 2018).

Para el desarrollo de nuestro modelo realizaremos un pre-proceso:

- Lectura de los datos mediante librería de Python pandas, dataframes.
- Normalización de la data.
- Generación de secuencias.
- Aplicación de LSTM cells.
- Flujo de entrenamiento y validación.
- Flujo de prueba.

Figura 2.4.2.11. Flujo de entrenamiento y validación.



Fuente: Microsoft Deep Learning, 2019.

CAPÍTULO 3

ANALISIS PRELIMINAR DE RESULTADOS

Antes de realizar modelos ya sea por análisis de series de tiempo mediante modelos ARIMA o por medio de modelos preliminares de machine learning, redes neuronales o LSTM, es necesario que se realice un preprocesamiento y análisis exploratorio de los datos obtenidos.

3.1. Procesamiento previo.

Acorde a las técnicas de ciencia de datos es necesario (U. San Diego, programa de ciencia de datos, 2018), para el buen desempeño del modelado estadístico, es necesario exista un procesamiento previo de la data, así como de un depurado de los misma.

La data obtenida corresponde a los precios históricos de oro, donde tenemos un periodo de tiempo que inicia desde el 27 diciembre de 1979 hasta el 1 de agosto del 2018, siendo estos datos diarios, es decir, acorde al calendario de apertura de mercados financieros de las commodities, esto nos da un primer cuadro de datos de 9772 filas (días) y con 5 columnas etiquetadas como fecha, precio, open, high y Low; siendo open el precio de apertura del oro, el high el precio más alcanzado en ese día y Low el precio más bajo alcanzado en ese día; cabe mencionar que esta data está en archivo .csv.

Adicional a esta data se obtuvo otra base de datos donde se encuentran los valores históricos de factores económicos que podrían tener cierta influencia en la definición del precio del oro como son el índice de inflación, Dow Jones, precio del petróleo Brent, la tasa de interés de los Estados Unidos, el índice del dólar y la producción mundial de oro. En esta base de datos tenemos los registros históricos desde el 2007 hasta el 24 octubre del 2017 dando un marco de datos de 2713 filas con 7 columnas (se incluye columna de la fecha) donde también el archivo se encuentra en .csv.

3.2. Limpieza de datos.

En este punto procederemos a eliminar la data que no nos ayude, revisar datos no concordantes y adicionar data faltante (espacios en blanco) para el análisis y posterior modelado.

La base de datos obtenida de los precios del oro y la data de cada uno de los factores macroeconómicos considerados tienen sus registros históricos con fechas de inicio diferentes, así también, las frecuencias de registro de los datos son distintas, por lo que, deberá considerarse en el análisis realizar los ajustes necesarios de la data frame para trabajar con las datas en la misma temporalidad. Adicional es necesario tener presente los escenarios macroeconómicos de los Estados Unidos, principalmente considerar las etapas de recesión económica, desde 1980 hubo 5 periodos de recesión económica, pero fácilmente agrupables en 3 siendo en el periodo de diciembre del 2007 a junio del 2009 la llamada gran recesión.

Tabla 3.2.1. Periodos históricos de recesión económica.

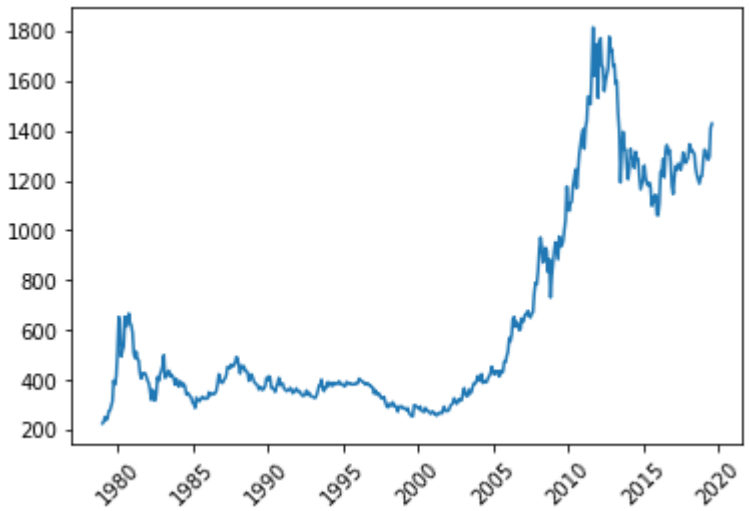
Nombre	Periodo	Duración (meses)	Tiempo desde la última recesión (meses)	Punta de desempleo	GDP (picos)
Recesión de 1980	Enero 1980-Julio 1980	6	4 años y 10 meses	7.8% (Julio 1980)	-2.2%
Recesión 1981-1982	Julio 1981-noviembre 1982	1 año y 4 meses	12	10.8% (Nov 1982)	-2.7%
Recesión temprana de 1990	Julio 1990-marzo 1991	8	7 años y 8 meses	7.8% (junio 1992)	-1.4%
Recesión temprana del 2000	Mar 2001-Nov 2001	8	10 años	6.3% (junio 2003)	-0.3%
La Gran Recesión	Dic 2007-junio 2009	1 año y 6 meses	6 años y 1 mes	10 % (octubre 2009)	-5.1%

Fuente: https://en.wikipedia.org/wiki/List_of_recessions_in_the_United_States, 2019

3.3. Caracterización de la muestra.

El siguiente gráfico muestra el desarrollo del precio del oro en el mercado internacional por onza, por un periodo de tiempo de 39 años desde 1970 hasta el 2018 (UC San Diego & Learning Python, ciencia de datos mediante Python, 2017).

Figura 3.3.1. Evolución histórica del precio internacional del oro



Fuente: El autor, datos portal de inversiones investing.com, 2020

A continuación, se describen los primeros 5 y los últimos 5 datos históricos del precio del oro, con su precio del día, precio de apertura, su precio más alto y el más bajo.

Figura 3.3.2. Precios iniciales diarios del oro de la base de datos

Date	usd
1978-12-31	226.0
1979-01-31	233.7
1979-02-28	251.3
1979-03-30	240.1
1979-04-30	245.3
1979-05-31	274.6
1979-06-29	277.5

Fuente: Portal de inversiones, investing.com, 2019

Figura 3.3.3. Precios finales diarios del oro de la base de datos

Date	usd
2019-01-31	1323.3
2019-02-28	1319.2
2019-03-29	1295.4
2019-04-30	1282.3
2019-05-31	1295.6
2019-06-28	1409.0
2019-07-31	1427.6

Fuente: Portal de inversiones, investing.com, 2019.

Se detalla la descripción estadística, de cada una de las columnas de la data.

Figura 3.3.4. Descripción estadística de columna de precio de oro.

	usd
count	488.000000
mean	642.029713
std	423.604866
min	226.000000
25%	348.275000
50%	406.650000
75%	952.875000
max	1813.500000

Fuente: El autor, 2019

Por la descripción estadística de los datos es posible ver que la media del precio del oro está en 642,03 dólares (lo que correspondería actualmente a 0,44 onzas de oro)

debido a que desde el año 1980 hasta aproximadamente el 2005, 25 años de data, los valores del precio del metal se mantuvieron en menos de los 600 dólares, esto es posible apreciarlo en los cuartiles de la data donde se manifiesta que 50% de la data tiene valores hasta los 407,9 dólares. A partir de los años 2005-2007 los valores del metal comienzan a tener una tendencia diferente, experimentando un alza casi sin paradas hasta el año 2011, día 22 de agosto donde se obtiene el máximo de 1888,7 dólares por onza de oro, a partir de este

punto el precio ha sufrido un descenso hasta un tope mínimo parcial o local de casi 1050,8 dólares la onza un 17 de diciembre del 2015.

La otra base de datos está compuesta por el histórico de posibles factores económicos externos que podrían tener alguna inferencia con el precio internacional del oro, esto para determinar algún modelo relacional mediante modelos regresivos multivariados.

Figura 3.3.5. Registro de datos históricos factores económicos externos.

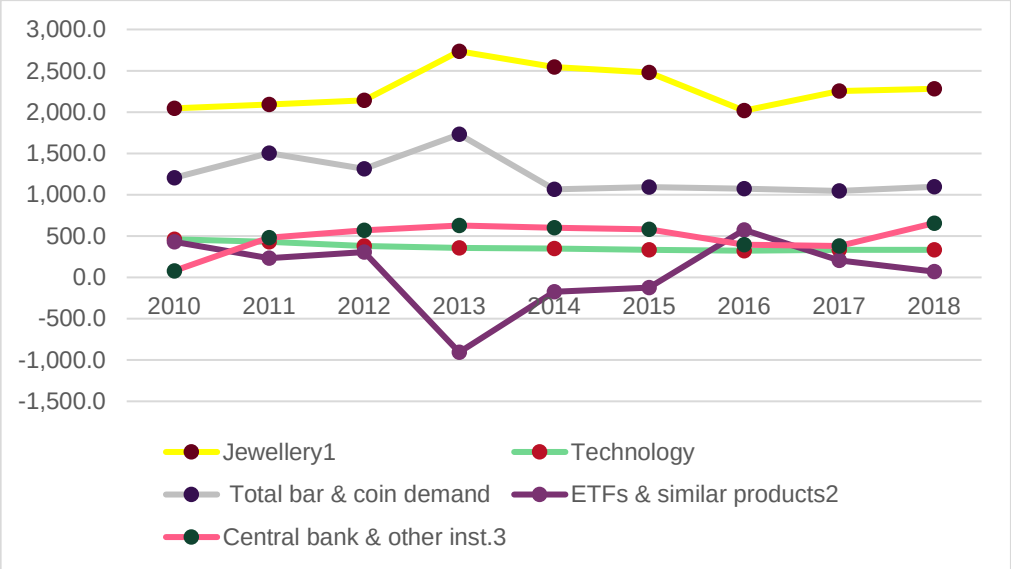
Date	usd	Inflat	dj	bo	inter	di
1/2/2007	652.55	2.852672	12673.68	56.72	5.248971	84.46
2/2/2007	654.75	2.852672	12653.49	58.41	5.248971	84.79
5/2/2007	648.00	2.852672	12661.74	58.10	5.248971	84.92
6/2/2007	653.90	2.852672	12666.31	58.42	5.248971	84.64
7/2/2007	655.25	2.852672	12666.87	57.23	5.248971	84.61

Fuente: El autor, 2019.

Observamos los primeros datos donde hemos incluido el precio del oro correspondiente a la fecha registrada.

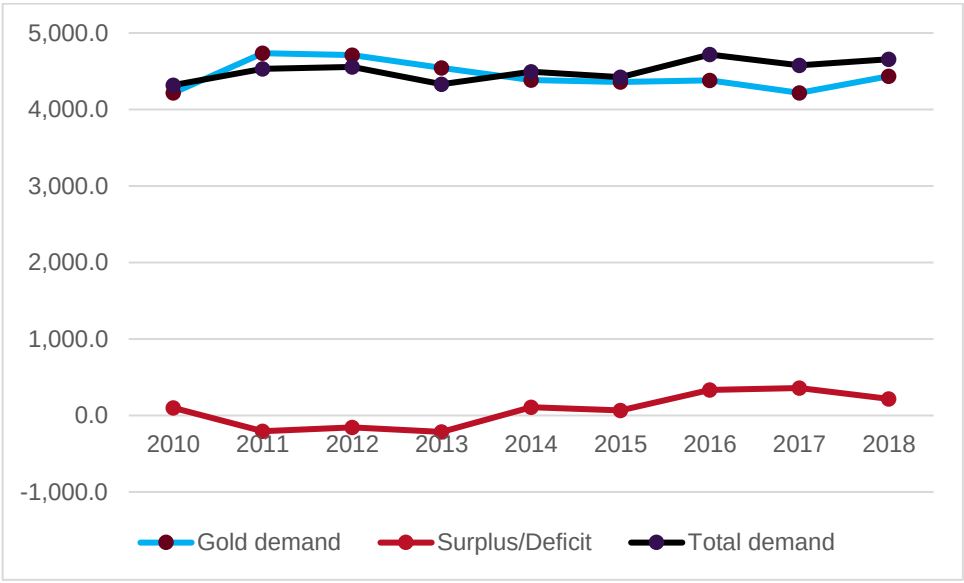
Hemos tomado los datos desde el 2007 por dos razones, primero porque desde el 2007 el precio del oro experimentó un cambio en su comportamiento casi constante de años anteriores, coincidentemente en los años 2007-2009 fue lo que se llamó la gran recesión por medio de la burbuja inmobiliaria y el oro es una fuente de refugio de las inversiones, por temas de oferta y demanda al aumentar la demanda de compra de oro.

Figura 3.3.6. Demanda mundial de oro desde el año 2010 al 2018 según la industria



Fuente: El autor, World Gold Council, 2019.

Figura 3.3.7. Demanda total y déficits de oro desde el año 2010 al 2018.



Fuente: El autor, World Gold Council, 2019.

Es posible observar como la demanda de oro llega a su punto máximo en el año 2011, punto donde se halla el precio máximo alcanzado por el oro, a mayor

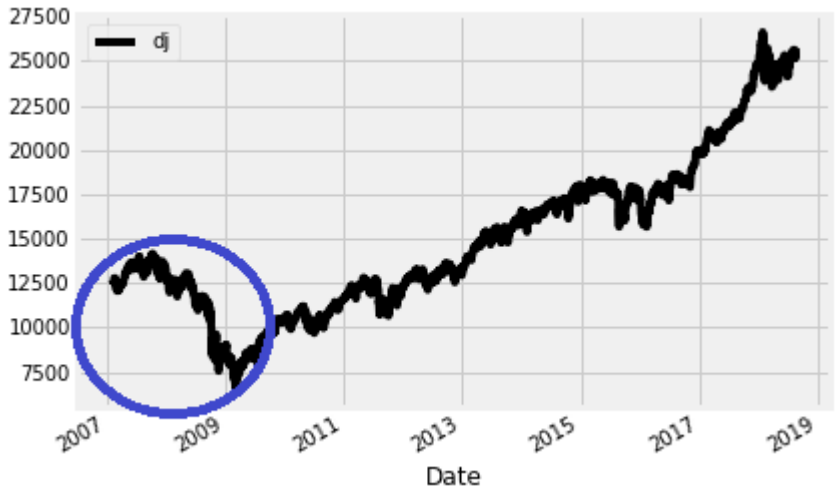
demanda los precios suben. Es posible observar cierta semejanza en las curvas de demanda de oro y precio del oro.

3.4 Gráficos temporales.

Se verá el comportamiento en el tiempo de cada una de las variables propuestas tanto los features o características como son la tasa de interés, índice de inflación, índice dólar, precio del barril del petróleo Brent y el índice Dow Jones, así como, la variable 'objetivo' que es el precio del oro. Estos gráficos serán desarrollados en la plataforma Python.

Índice Dow Jones

Figura 3.4.1. desarrollo histórico del precio del Dow Jones.

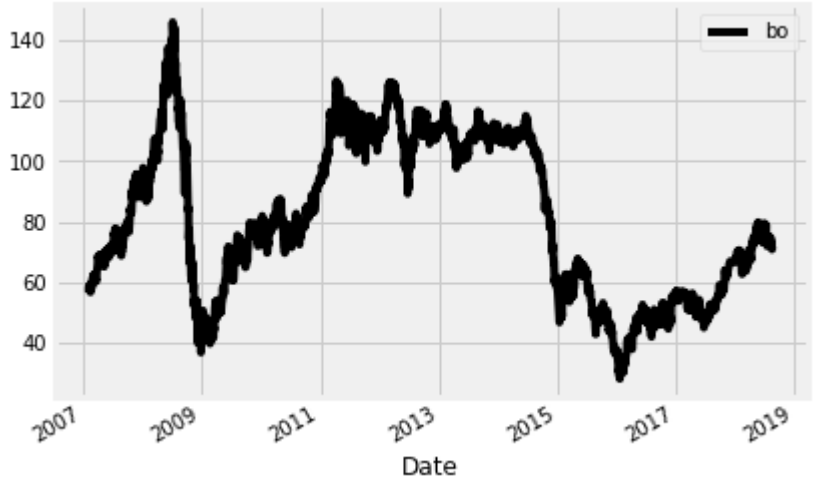


Fuente: macrotrends.net, 2019

Encerrado con circulo es el periodo desde el 2007-2009 de la gran recesión.

Brent Oil

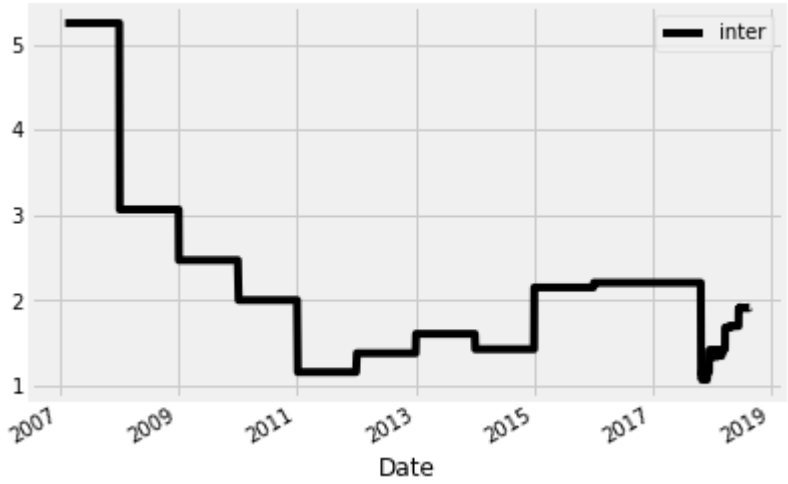
Figura 3.4.2. Desarrollo del precio internacional del barril de petróleo tipo Brent.



Fuente: Inversiones, investing.com, 2019.

Tasas de interés de USA

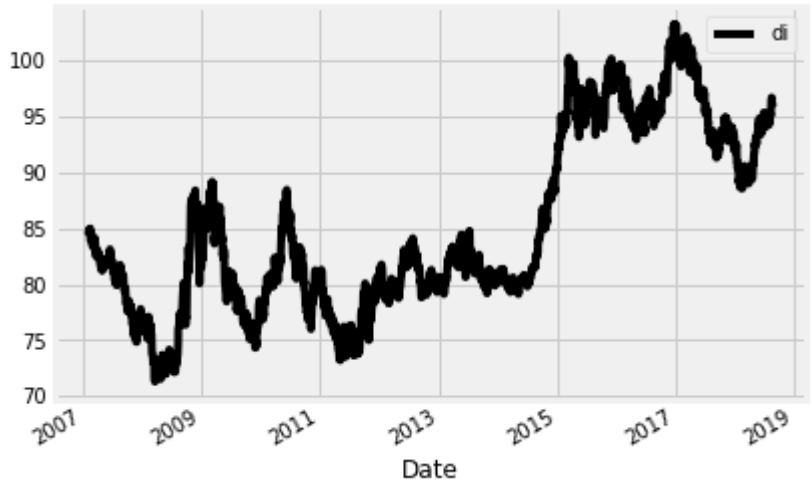
Figura 3.4.3. Desarrollo histórico de las tasas de interés de E.E.U.U.



Fuente: Macrotrend.net, 2019.

Dólar Índex

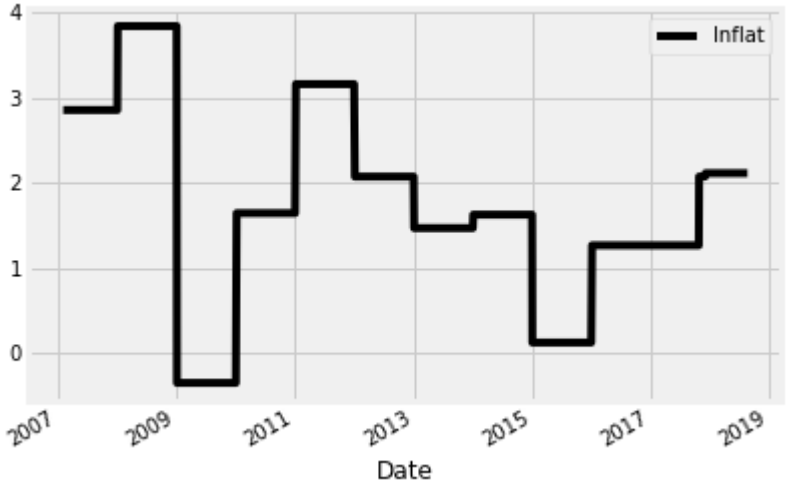
Figura 3.4.4. Desarrollo histórico del índice de dólar.



Fuente: Macrotrend.net, 2019.

Tasas de inflación

Figura 3.4.5. Desarrollo histórico de las tasas de inflación de los E.E.U.U.

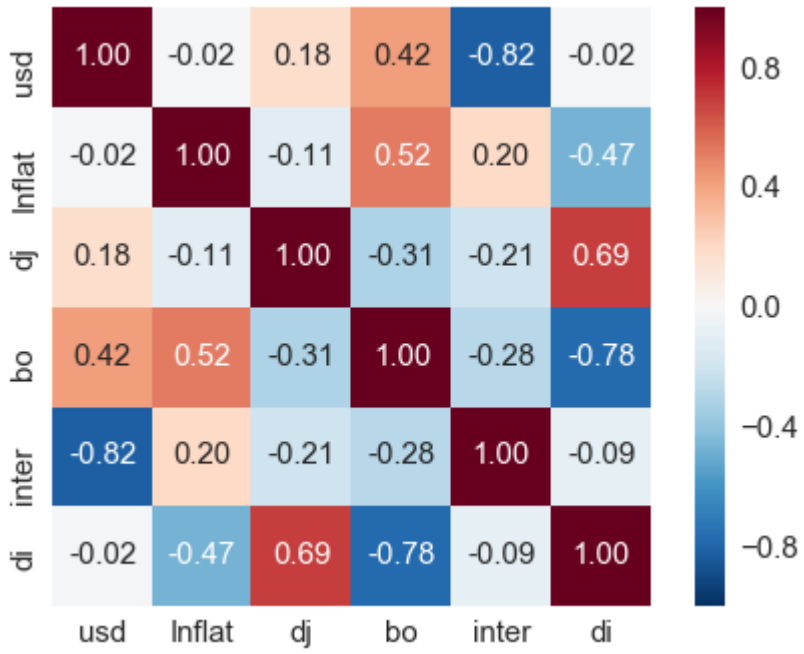


Fuente: Macrotrend.net; 2019.

Una vez obtenidos los desarrollos temporales de cada una de las variables es posible empezar a desarrollar modelos univariados y multivariados.

Se han calculado el índice de correlación entre cada una de las variables tanto los “features” como la variable “objetivo”, lo cual hemos desarrollado mediante un mapa de calor teniendo los siguientes resultados (Mark Lutz, learning Python, 2015).

Figura 3.4.6. Mapa de calor de los índices de correlación de cada variable y el precio del oro



Fuente: El autor, 2019.

Donde preliminarmente se puede manifestar que existe una correlación alta negativa con las tasas de interés (-0.82) seguido del índice de correlación con el precio del Brent oíl (0.42).

CAPÍTULO 4

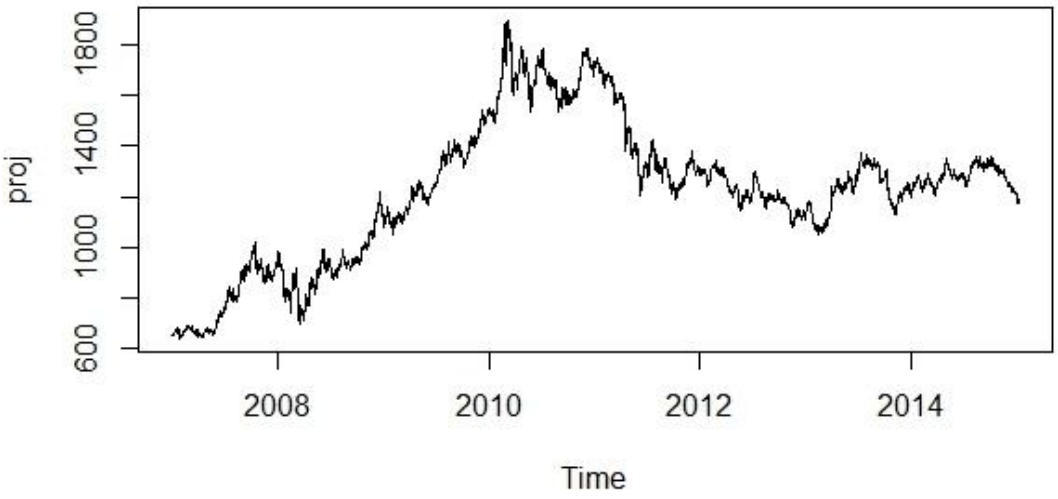
ANALISIS COMPLETO DE DATOS Y RESULTADOS

4.1 Modelo ARIMA.

Como ya se mencionó los datos fueron ordenados en precios promedio mensuales, figura #29. Se observo la serie de tiempo y no presenta una estacionalidad, es decir, no tiene una media fija y una varianza uniforme, por lo que fue necesario realizar una diferenciación y una transformación logarítmica para quitar el crecimiento exponencial, luego se graficó el factor de auto correlación ACF y el factor de auto correlación parcial PACF (Francisco Vera, PhD, notas de clases, 2017).

En el gráfico de serie de tiempo se observaron tres zonas, una de crecimiento hasta límite superior máximo alcanzado en el 2011 (crecimiento sostenido), una zona de decrecimiento hasta aproximadamente el 2012 y una zona de tendencia constante o estancamiento aproximadamente desde el 2012 en adelante como se puede observar en la figura#28.

Figura 4.1.1 Desarrollo del precio del oro desde el 2007 hasta 2018.



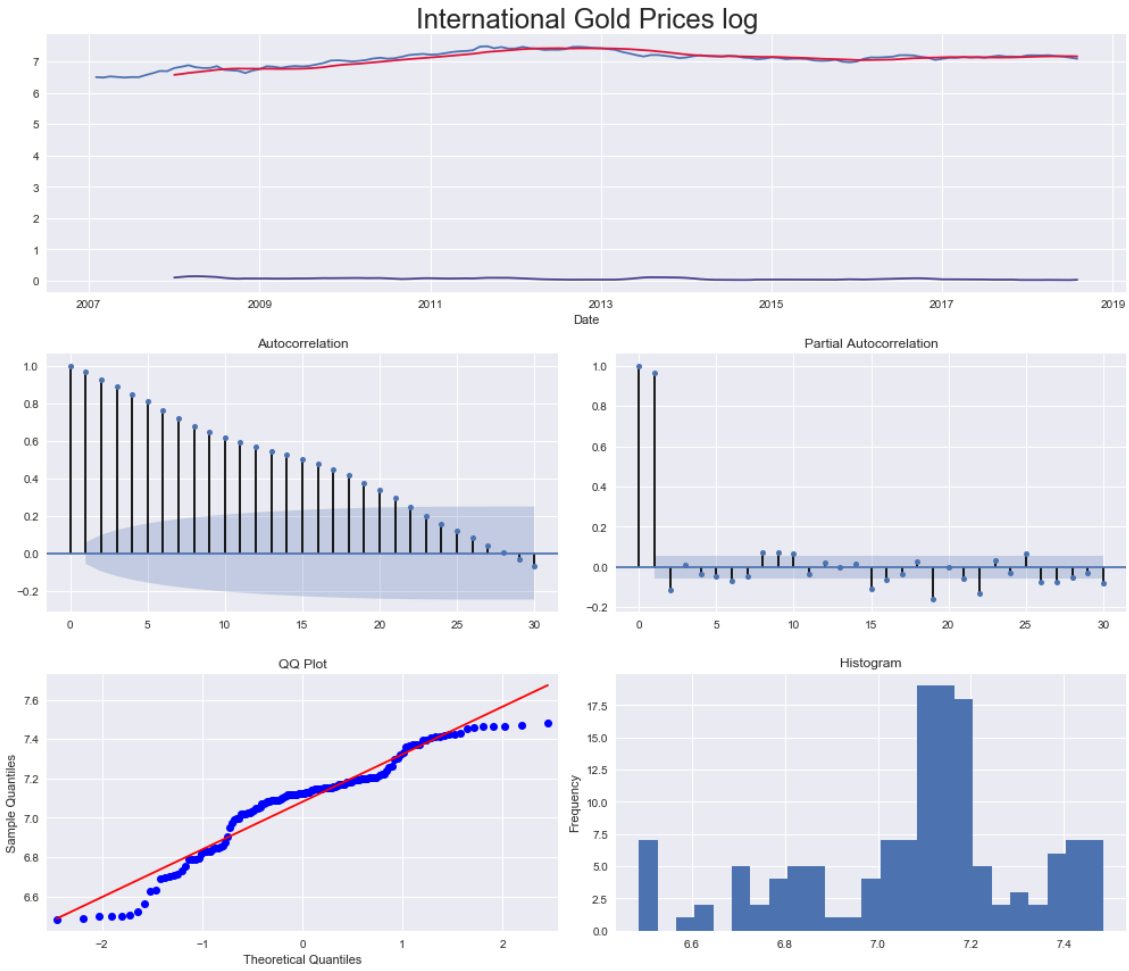
Fuente: El autor, 2019.

Figura 4.1.2. Desarrollo del Precio del oro / mes.

Fuente: El autor, 2019.

No fue necesario aplicar una transformación logarítmica de los datos ya que estos no presentaron crecimientos exponenciales aparentes, más, se realizó la trasformación logarítmica de los datos para verificación, obteniendo las siguientes observaciones, figura 4.1.3 (Francisco Vera, PhD, notas de clases, 2017 y Daniel Peña, A course of time series análisis, 2001).

Figura 4.1.3. Prueba de estacionalidad de los datos.



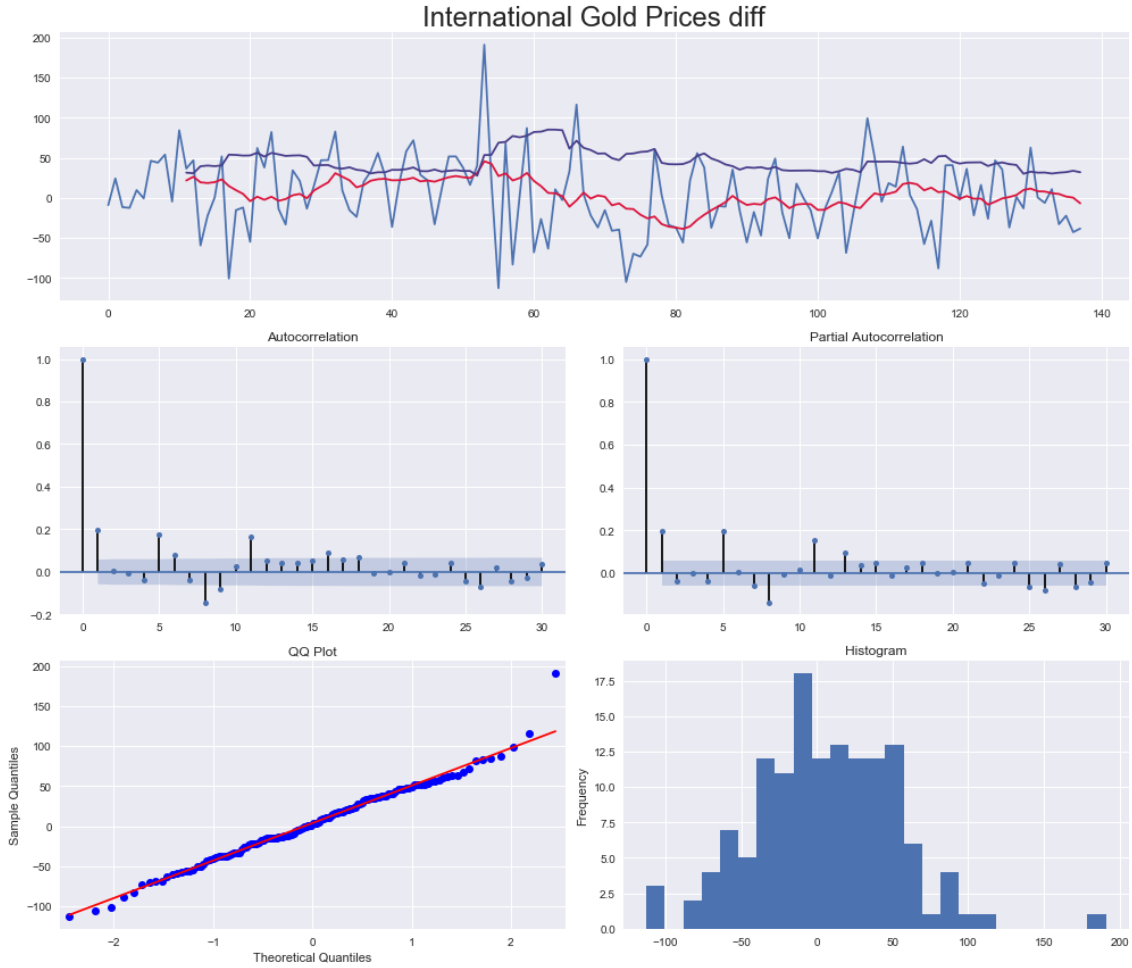
Resultados Dickey-Fuller Test:	
Test estadistico	-2.587227
p-value	0.095666
# de lags	1.00
# de observaciones	137.00
Valor critico (1%)	-3.479007
Valor critico (5%)	-2.882878
valor critico (10%)	-2.578149

Fuente: El autor, 2019.

Se observó las tendencias de los datos y un desarrollo de la desviación estándar constante. El valor P presenta un valor alto 0.09 superior al 0.05 considerando la hipótesis con un intervalo de confianza del 95%. Los gráficos de ACF y PACF presentan residuales no aleatorios y los datos no se ajustan a recta de la figura Q-Q, no ajustándose a la distribución normal.

Se realizó una primera diferenciación de los datos obteniendo la siguiente información, figura 4.1.4.

Figura 4.1.4. Prueba de estacionalidad de los datos, primera diferenciación.



Resultados Dickey-Fuller Test:	
Test estadistico	-9.46E+00
p-value	4.28E-16
# de lags	0.00E+00
# de observaciones	137
Valor critico (1%)	-3.479007
Valor critico (5%)	-2.882878
valor critico (10%)	-2.57814

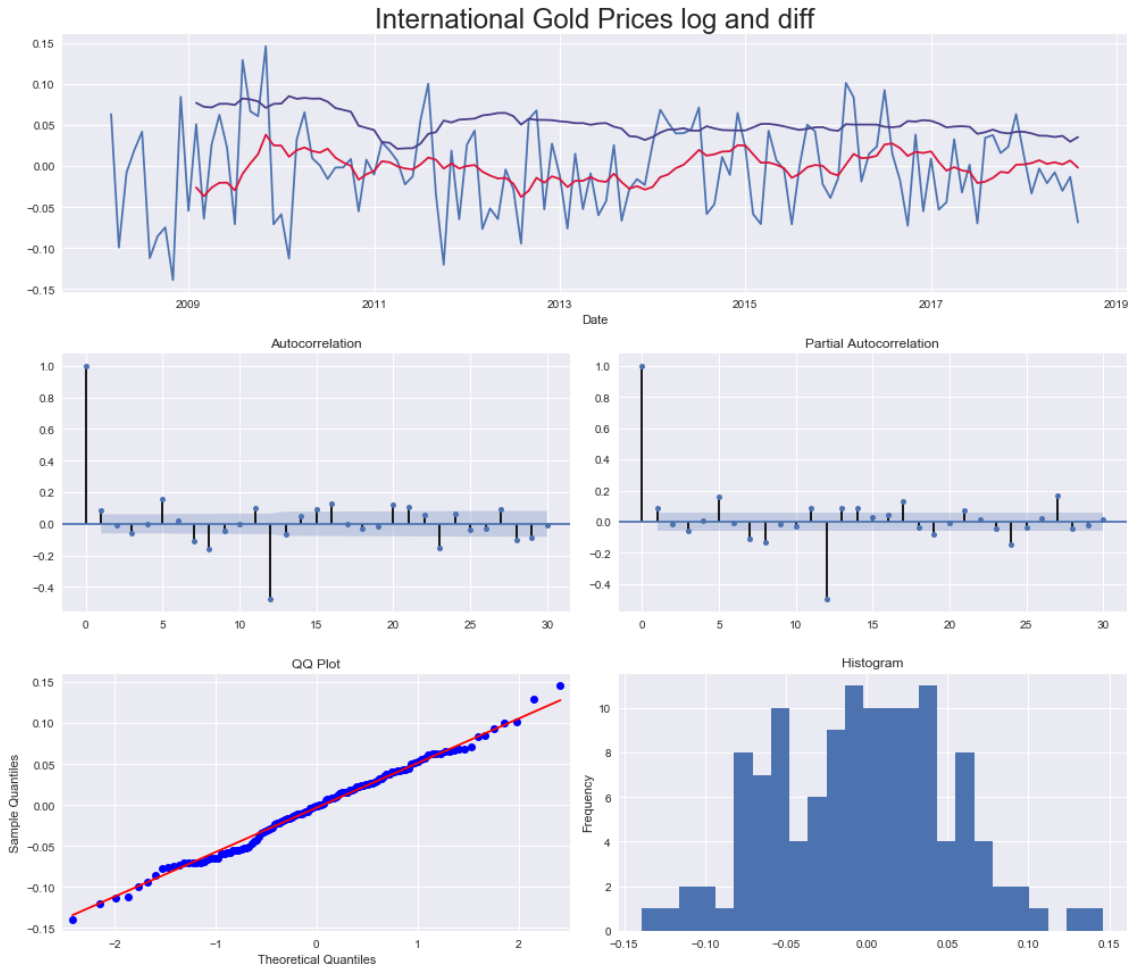
Fuente: El autor, 2019.

Se observo un ajuste de los datos en comparación a la distribución normal, pero aún se observan en los gráficos ACF y PACF ciertas correlaciones, picos sobresalientes de los residuos, a pesar de que el test del estadístico es mucho menor que 0.05 y el p-value mantiene un valor más bajo que 0.05 tomado como

nivel de significancia. Es posible observar que existe una estacionalidad de los datos al observar picos sobresalientes en un lag o frecuencia de 12.

Se realizará una segunda transformación para observar el comportamiento.

Figura 4.1.5. Prueba de estacionalidad de los datos diferenciados y transformados logarítmicamente.



Resultados Dickey-Fuller Test:	
Test estadistico	-4.441800
p-value	0.000250
# de lags	12.00
# de observaciones	137.00
Valor critico (1%)	-3.489600
Valor critico (5%)	-2.887500
valor critico (10%)	-2.580600

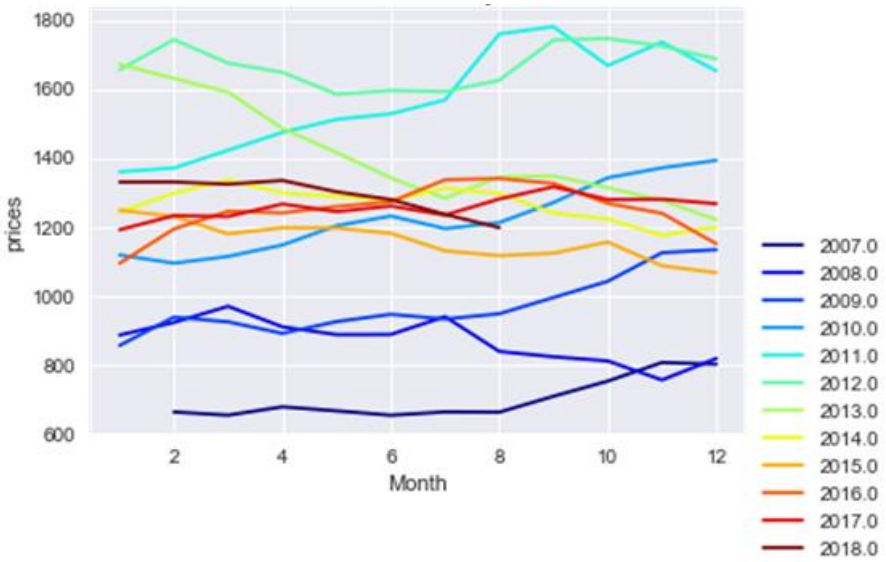
Fuente: El autor, 2019.

Fue posible observar que en la primera diferenciación hay un mejor ajuste de los datos a la figura del Q-Q, así también el estadístico es más alto que el obtenido,

en la primera diferenciación. El test del estadístico es mucho menor que el p-value y este a la vez mantiene un valor más bajo que 0.05 tomado como nivel de significancia

Se revisará la estacionalidad observando un gráfico de los precios de oro por mes y por año, figura #33 (Mark Lutz, Learning Python, 2013).

Figura 4.1.6. Comportamiento del precio del oro por mes.



Fuente: El autor, 2019.

Es posible observar una tendencia de los precios a descender aproximadamente a partir del octavo mes, existe una excepción de los datos a aumentar los años 2007 al 2009, a partir del octavo mes, esto quizás debido a la gran recesión producida justamente durante estos años, donde se observa este comportamiento ascendente.

El modelo ARIMA estacional con el menor AIC encontrado, desarrollando modelo de iteración de factores de auto correlación de movimiento promedio y estacional, es el siguiente

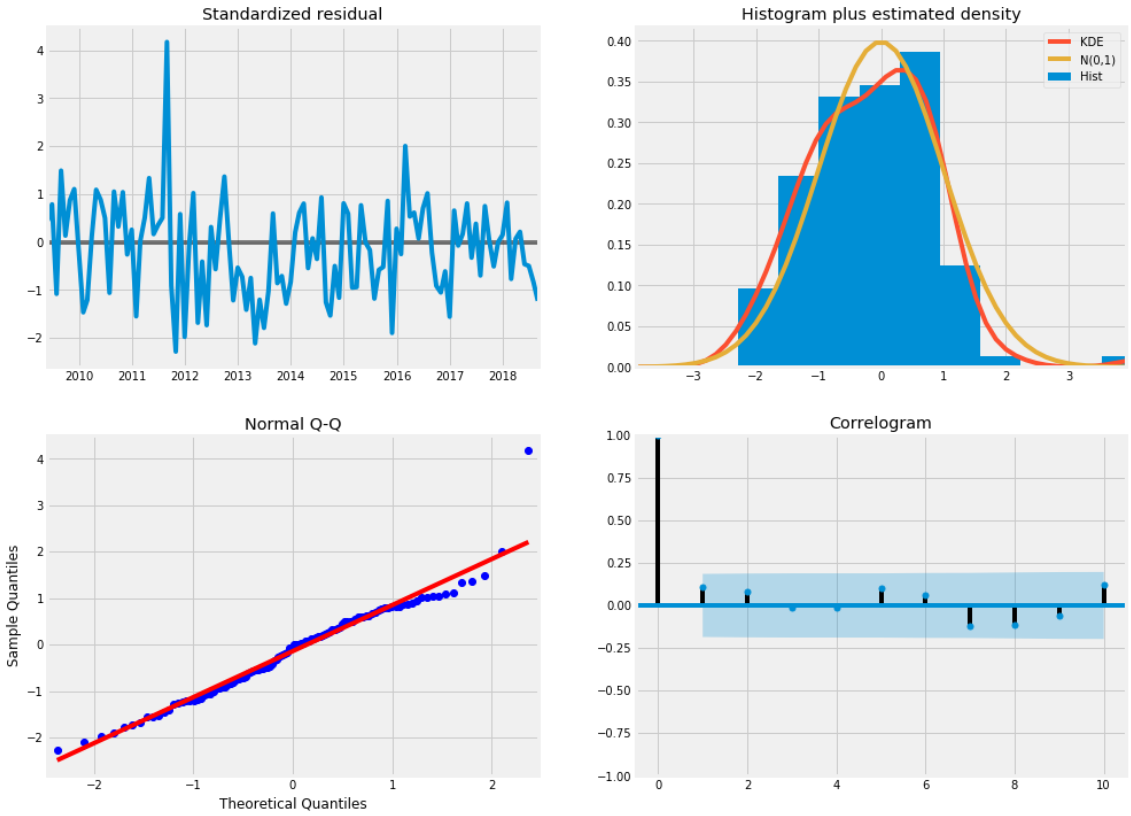
ARIMA (1, 1, 1) x (0, 1, 1, 12)12 - AIC:1201.637513

Seguido por,

ARIMA (1, 1, 1) x (1, 1, 1, 12)12 - AIC:1203.331711

Graficando el resumen estadístico del modelo:

Figura 4.1.7. Resumen Estadístico modelo ARIMA (1,1,1) (0,1,1,12)12.



Fuente: El autor, 2019.

Se observó una independencia de los residuos mediante el correlograma (baja correlación con otras versiones con lags), como una variancia y media alrededor de cero, un ajuste muy bueno de los datos a la gráfica Q-Q a la distribución normal, así como una buena aproximación a del KDE a la recta N(0,1), es decir, una buena indicación que los residuos están normalmente distribuidos.

Al haber realizado el modelo predictivo tomado desde el 2017 hasta el 2018-31-08 obtuvimos el siguiente gráfico, figura#35

Figura 4.1.8. Prueba de modelo predictivo desde 2017 hasta agosto del 2018.

Fuente: El autor, 2019.

Lo que nos dio una calidad de predicción en MSE y RMSE, de:

MSE = 740.72 MSE (27.22 RMSE)

Con nueva data tenemos la siguiente predicción 6 meses después,

Figura 4.1.9. Predicción en un periodo de 6 meses adelante.

Fecha	Precio (usd)
9/30/2018	1213.49
10/31/2018	1213.30
11/30/2018	1205.95
12/31/2018	1193.26
1/31/2019	1213.74
2/28/2019	1246.45

Fuente: El autor, 2019.

4.2 Modelo relacional.

Para el modelo relacional, se tomaron 5 variables macroeconómicas externas, para determinar su relación con el precio internacional del oro, mediante tabla de datos mensuales, comprendidos desde el 2007 hasta agosto del 2018, se detalla la muestra en la figura 43 y 44, datos iniciales y finales.

Figura 4.2.1 Datos iniciales de muestra de factores económicos externos.

Date	usd	Inflat	dj	bo	inter	di
2007-02-28	665.102500	2.852672	12638.758750	58.804000	5.248971	84.273000
2007-03-31	655.890909	2.852672	12268.533636	62.455000	5.248971	83.343182
2007-04-30	680.007895	2.852672	12764.642105	67.651579	5.248971	81.895263
2007-05-31	668.309524	2.852672	13412.269524	67.935238	5.248971	82.029524
2007-06-30	655.714286	2.852672	13480.212857	70.538571	5.248971	82.360000
2007-07-31	665.265909	2.852672	13672.801364	75.817727	5.248971	80.633636
2007-08-31	664.529545	2.852672	13235.965909	71.256818	5.248971	80.877273
2007-09-30	710.645000	2.852672	13549.967000	76.963000	5.248971	79.312000
2007-10-31	754.480435	2.852672	13901.280435	82.475217	5.248971	77.779565
2007-11-30	808.311364	2.852672	13186.458636	92.204091	5.248971	75.629318

Fuente: El autor, 2019.

Figura 4.2.3. Datos finales de muestra de factores económicos externos.

Date	usd	Inflat	dj	bo	inter	di
2017-11-30	1281.189091	2.070000	23557.209545	64.144545	1.155909	93.961818
2017-12-31	1268.036364	2.108182	24564.974545	66.870000	1.293636	93.331818
2018-01-31	1330.600000	2.110000	25760.009130	68.982609	1.412174	90.820870
2018-02-28	1330.945000	2.110000	24987.071500	65.730500	1.416000	89.694500
2018-03-31	1324.611364	2.110000	24549.965000	66.821364	1.490909	89.876818
2018-04-30	1335.342857	2.110000	24304.212857	71.762381	1.691429	90.293333
2018-05-31	1302.400000	2.110000	24571.868696	77.006522	1.699565	93.336522
2018-06-30	1280.023810	2.110000	24790.108095	75.941429	1.810952	94.371905
2018-07-31	1237.145455	2.110000	24945.847273	74.951818	1.910000	94.608636
2018-08-31	1198.709375	2.110000	25454.056875	72.504375	1.911875	95.836250

Fuente: El autor, 2019.

La simbología de cada uno de los factores económicos se detalla tal como, precio internacional del oro (USD), tasa de inflación norteamericana (Inflat), índice Dow Jones (dj), precio internacional del barril de petróleo tipo Brent (bo), tasa de interés norteamericano (inter) y el índice internacional del dólar (di).

Se estableció un modelo de regresión lineal múltiple de la forma:

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon$$

(4.1)

$\beta_{0,..,k} = Coeficentes$

$X_{1i,..,ki} = Predictores, variales\ independientes.$

$\varepsilon = Error\ aleatorio$

Este modelo se realizó con Python 3.6 usando módulo estadístico “Regresión lineal” para modelos de regresión multivariable.

Realizado este algoritmo obtenemos el siguiente modelo resultante.

Figura 4.2.4. Resumen estadistico de modelo relacional multivariable en Python.

Dep. Variable:	y	**R-squared:**	0.745
Model:	OLS	**Adj. R-squared:**	0.739
Method:	Least Squares	**F-statistic:**	131.6
Date:	Wed, 26 Sep 2018	**Prob (F-statistic):**	6.79e-40
Time:	23:21:21	**Log-Likelihood:**	-883.07
No. Observations:	139	**AIC:**	1774.
Df Residuals:	135	**BIC:**	1786.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	424.1278	314.591	1.348	0.180	-198.036	1046.292
x1	4.7578	0.880	5.405	0.000	3.017	6.499
x2	-179.4383	13.924	-12.887	0.000	-206.976	-151.900
x3	9.4627	2.753	3.437	0.001	4.018	14.907

Omnibus:	4.699	**Durbin-Watson:**	0.234
Prob(Omnibus):	0.095	**Jarque-Bera (JB):**	2.866
Skew:	0.143	**Prob(JB):**	0.239
Kurtosis:	2.357	**Cond. No.**	3.12e+03

Fuente: El autor, 2019.

A este punto se eliminaron dos de las variables independientes (factores macroeconómicos), Inflación de los Estados Unidos (Inflat) y Dow Jones index (dj), ya que, los valores de contraste de hipótesis nula (p-value) eran mayores al nivel de significación escogido ($p \leq 0{,}05$), y así obtener un resultado estadísticamente significativo. El módulo y método OLS se obtiene la minimización de los cuadrados de las distancias de la línea de regresión.

El valor correspondiente a R-Square es el porcentaje de varianza explicada por el modelo, es decir, donde la varianza del error es menor que la varianza de y. En nuestro modelo el R-square es de 0.745, lo que quiere decir que este modelo explica el 74,5% de la varianza en nuestra variable dependiente

Los predictores que resultaron relacionados fueron el precio del petróleo tipo Brent (x1), la tasa de interés *(x2)* y el índice dólar *(x3).* Se obtuvo la siguiente ecuación:

$$E(Y|X_1, X_2, X_3)=424.1278 + 4.7578(X_1) - 179.4383(X_2) + 9.4627(X_3);$$

(4.2)

Y = Precio del oro estimado.

X1= precio internacional de petróleo tipo Brent.

X2= Tasa de interés de los estados Unidos.

X3=Índice del dólar.

4.3 Modelo de redes neuronales.

En la figura #38, se puede observar parte de la codificación en Python para la conformación de las dos capas 10 y 300 neuronas.

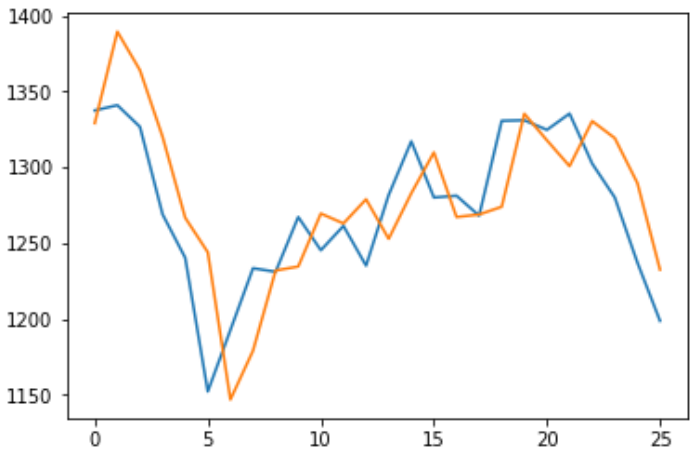
Figura 4.3.1. Aplicación de célula LSTM, Python

```
model = Sequential()
model.add(LSTM(input_shape = (10,1), output_dim= 10, return_sequences = True))
model.add(Dropout(0.5))
model.add(LSTM(300))
model.add(Dropout(0.5))
model.add(Dense(1))
model.add(Activation("linear"))
model.compile(loss="mse", optimizer="adam")
model.summary()
```

Fuente: El autor, 2019.

La paramétrica usada para medir el error es el error cuadrático entre (z, Y).
Se ha realizado una red neuronal de dos capas, con 10 neuronas en una y 300 en la segunda, obtenido los siguientes resultados.

Figura 4.3.2. Predicción de prueba de datos conocidos, LSTM.



Fuente: El autor, 2019.

MSE = 1459.3280

MAE= 31.9335

RMSE= 38.2011

Tabla 4.3.1. Comparativa de precios reales versus los pronosticados.

Precios Reales	Precios Pronosticados
1335.34	1300.69
1302.40	1330.39
1280.02	1319.16

1237.14	1289.55
1198.70	1232.49

Fuente: El autor, 2019.

Al realizar la comparación de las métricas de los tres métodos aplicados a la predicción del precio del oro se resumió lo siguiente:

Tabla 4.3.2. Métricas de comparación entre métodos estadísticos predictivos de series de tiempo.

Parámetro	Regresión Lineal	SARIMA	LSTM
Modelo	$E(Y\|X_1, X_2, X_3)$=424.1278+4.7578(bo)-179.4383(inter)+9.4627(di)	(1,1,1) (0,1,1,12)	Dos capas (10 x 300) neuronas
AIC	1774	1201.63	NA
MSE	19304.6124	740.72	1459
MAE	117.6249	21.02	31.93
RMSE	138.9410	27.22	38.20

Fuente: El autor, 2019.

CAPÍTULO 5

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones.

1.-Se analizó la estadística descriptiva de la evolución del precio del Oro, se encontró registros de los precios diarios del oro desde el año 1970 hasta agosto del 2018. Para nuestro análisis consideramos los precios como medias mensuales desde el 2007 hasta agosto 2018, dándonos una data de 2929 filas de datos y una media de 1222.6 USD. Para realizar este proyecto hubo que realizar un ajuste de los datos de los diferentes factores macroeconómicos usados para que todos estén bajo los mismos periodos de tiempo y cantidades. La labor de buscar, ordenar y limpiar generalmente es la actividad que mayor tiempo consume en las actividades de ciencia de datos.

2.-Dentro de los 5 factores para determinar una relación lineal con el precio del oro tales como el índice de inflación, índice del Dow Jones, precio internacional del petróleo tipo Brent, las tasas de interés de los estados Unidos y índice del dólar, se encontró una relación mediante un modelo de regresión lineal con solo tres de los 5 factores, estos son precio de petróleo tipo Brent, tasas de interés y el índice dólar.

3.-Se determinó un modelo ARIMA estacional para pronóstico mensual del precio del oro, siendo el modelo con el más bajo *AIC =1201.63*, SARIMA *(1,1,1) (0,1,1,12).*

4.- Se aplicó un modelo básico de redes neuronales recurrentes RNN con células del tipo LSTM, el cual nos dio una medida del error muy cercano al obtenido al modelo ARIMA, esta diferencia radica por la cantidad de data trabajada, las variaciones de factores del modelo neuronal y los tiempos computacionales elevados a equipos de cómputo “no especializados”.

5.- Acorde a las métricas se encuentra que el modelo SARIMA (1,1,1) (0,1,1,12), es modelo con menor RMSE en comparación a los tres métodos aplicados tabla 4, por lo que nos dará los mejores pronósticos del precio del oro.

Si bien es cierto las métricas de error son mejores con el desarrollo de un modelo SARIMA, es necesario manifestar que las técnicas de redes neuronales pueden ser optimizadas, dependiendo de la experiencia del investigador realizando modificaciones a la estructura del modelo, es decir, realizar modificaciones a ciertos factores como son el número de capas, numero de perceptrones, capacidad del CPU, numero de batchs y de epochs para iteración de aprendizaje. Adicional se debe siempre considerar que estos modelos de redes neuronales, al usar los datos históricos como parte del input en cada time step, son modelos que trabajan de manera óptima con bases de datos grandes. En nuestro caso de estudio se usaron como datos los precios promedio mensuales desde el 2007 hasta el 2018 dándonos una base de datos de 129 valores, lo cual se considera una base de datos pequeña.

También hay que considerar que no siempre un modelo de redes neuronales será el óptimo, pero los métodos que mejor funcionan son los modelos de ARIMA y los modelos redes neuronales recurrentes con LSTM para análisis de series de tiempo.

5.2. Recomendaciones.

- Generar modelos optimizados de redes neuronales aplicados a series de tiempo.
- Desarrollar a profundidad los conceptos matemáticos y computacionales de las redes neuronales y el aprendizaje de máquina aplicados a series de tiempo.
- Realizar el modelo neuronal considerando la data de precios diaria del oro, el cual llega a una tabla de 10600 filas, pero tiempos computacionales altos.

REFERENCIAS BIBLIOGRAFÍCAS

Ian Goodfellow, Yoshua Bengio, & Aaron Courville. (2016). *Deep Learning*, Massachusetts Institute of Technology.

Aurelien Geron. (2017). *Hands-on Machine Learning with SciKit and TensorFlow*. O'Reilly.

Jason Brownlee. (2018). *Deep Learning for time series Forecasting*. Machine Learning Mastery.

Kirill Eremenko, & Hadelin de Ponteves,. (2018). *Machine Learning A-Z: Hands-on Python & R in Data Science*.Curso Internacional Udemy.

Taegyun Jeon, D. (2016). *Recurrent Neural Networks, basics and Implementations*. Senior Resercher, R&D Center. Satrec Initiative.

Mark Lutz. (2013). *Learning Python*. O'Reilly.

Google. (2018). *Colab research*.

Francisco vera, PhD. (2017). *Pronóstico de la demanda*. Apuntes y videos de clases de series de tiempo.

Jiménez Daniela. (2011). *Análisis y pronósticos de demanda para telefonía móvil*. Tesis maestría en gestión de Operaciones, Universidad de Chile.

Larose, T. Daniel., Larose, D. Chantral., (2015). *Data Mining and predictive analytics*. John Wiley & sons.

Hossein Mombeini & Abdolreza yazdani. (2015). *Modeling gold Price via Artificial Neural Network*. Journal of economics, Business and management,vol 3 No7.

Banhi Guha & Gautam Bandyopadhyay. (2016). *Gold Price forecasting using ARIMA Model.* Journal of Advanced Management science, Vol 4 No2.

Dimitri Pissarenko. (2002). *Neural Networks for financial time series prediction*: Overview Over Recent Research. BSc Computer Studies.

Julian Faraway & Chris Chatfield. (1998). *Time series forecasting with neural networks: a comparative study using the airline data*. Applied, Stats 47 part 2 pp 231-250.

Daniel Peña, George Tiao & Ruey Tsay. (2001). *A course of time series Analysis*. Wiley series.

Hamideh Moradi, Iman Jokan & Ahmad Forouzantabar, (2015). *Modelling and forecasting gold price using GMDH neural network*. Indian Journal of fundamental and applied Life science ISSN: 2231-6345.

Simon Haykin. (2009). *Neural Networks and Learning Machines*. Pearson Education, Inc. Third Edition.

Sayan Pathak, (2017). ML scientist Microsoft tutorial, Deep Learning Explained. Tutorial.

Karpathy Andrej (2015), *Deep Neural Networks*, Blog, Stanford and director of AI of Tesla.

Datos históricos del oro (2019). www.Investing.com.

Oferta y demanda mundial del Oro (2018). World Gold Council. www.gold.org.

Datos de producción de oro en Ecuador (2018). Estadísticas del Ministerio de Minería del Ecuador.

Shahriar Shafiee & Erkan Topal, (2009). An overview of global gold market and gold price forecasting. Elsevier, global information analytics.

James Chen, Will Kenton & Julia Kagan, (2019). Investopedia.

José Ángel Fernández, (2003), *Técnicas cuantitativas elementales de previsión univariante*.