

ANÁLISIS DE SERIES DE TIEMPO PARA LA PREDICCIÓN DEL PRECIO DEL EUR/USD CON ARIMA

**Juan Camilo Serrano Correa
Rafael Fernando Leal Ramirez**

Forecasting, Series de tiempo

Introducción

Una serie de tiempo es una secuencia tomada en datos sucesivos en el tiempo discreto

Componentes que suelen tener estos datos:

Tendencia: Comportamiento relativamente constante a lo largo del tiempo

Cíclico: Fluctuación en forma de onda alrededor de la tendencia

Estacionario: Se repita anualmente, clima, agrícola

Aleatoriedad: variabilidad que queda después

cualquier variable que cambia con el tiempo, puede ser considerada para la aplicación que se propondrá a continuación. Algunos ejemplos de la vida diaria que aplican son:

Previsión del número de pacientes

Pronóstico de visitantes de un sitio web

Pronóstico de visitantes de un lugar

Proyecciones de ventas

pronóstico del clima

Con los ejemplos dados, es fácil imaginar la infinidad de aplicaciones y problemas que puede llegar a solucionar.

Descripción y delimitación del problema

Los pronósticos se realizan con el fin de orientar las decisiones que se van a tomar en un futuro determinado. Como la previsión del número de pacientes en un hospital, para así preparar el equipo que los atenderá, por ejemplo, distribuir los médicos, enfermeros, auxiliares que atenderán, alistar medicamentos e insumos. Pronóstico del clima, con importantes aportes a la agricultura, pronosticar la presencia del fenómeno del niño, etc.

Otra gran aplicación que puede tener es la predicción en el mercado de valores, ahí se pueden comprar y vender acciones de alguna empresa, materias primas e incluso divisas extranjeras. Por lo general se hace con el objetivo de aumentar el dinero invertido, pero diversas estadísticas exponen que más del 90% de las personas que invierten dinero en bolsa terminan con pérdidas. Si miramos con más detalle estas estadísticas nos damos cuenta que gran parte de estas personas son inexpertas, pero incluso para las personas

más expertas en este campo; invertir en la bolsa de valores representa un gran desafío e incertidumbre de cuál será el precio de determinado activo en el futuro.

Más específicamente, nos concentramos en la predicción del par de divisa EURO - DOLAR es el instrumento más negociado del mundo, esto se debe a diversas razones como ser la moneda de los países con las economías más grandes, lo que repercute en ser el par de divisa que al variar su precio a más personas afecta, como lo es en exportaciones e importaciones, apreciación o depreciación de otras monedas o particularmente a grandes bancos, empresas multinacionales, gobiernos y especuladores que acostumbran a negociar esta divisa.

Objetivo

Desarrollar una herramienta que permita pronosticar el comportamiento que tendrá el EUR/USD en el futuro.

Propuesta a desarrollar

Se propone un modelo que sea capaz de pronosticar el precio que tendrá el EUR/USD. Para el pronóstico del par de divisas EUR/USD, que son datos de series de tiempo, requiere de una estructura diferente de los datos para que pueda ser resuelto, mediante el modelo arima, a diferencia del mediante aprendizaje supervisado que no. La figura 1, expone claramente la principal diferencia entre los datos para series de tiempo y machine learning.

Machine learning data set →	Sensor ID	Value
	Sensor_1	20
	Sensor_1	21
	Sensor_2	22
	Sensor_2	23

Time series data set →	Sensor ID	Time Stamp	Value
	Sensor_1	01/01/2020	20
	Sensor_1	01/02/2020	21
	Sensor_2	01/01/2020	22
	Sensor_2	01/02/2020	23

[11]

En el pronóstico de las series de tiempo se trata sobre el modelo arima.

El concepto de par de divisas es un concepto esencial en finanzas. Su importancia radica en su cotización oficial. La cotización oficial de una divisa frente a otra se puede conocer en el mercado de divisas. [1] En este caso el par euro-dólar nos dice la cantidad de dólares a la que equivaldría la moneda base que es el euro.

El par euro-dólar tiene una relevancia grande para la economía por ser las divisas con países con las economías más grandes, por lo que su valoración o devaluación va a tener

un impacto significativo sobre el valor de otras divisas, por esto las grandes empresas como bancos o multinacionales suelen negociar en estas divisas.

Se puede observar que las divisas cambian de valor con el tiempo y pueden subir o bajar de un día a otro, por lo que se puede aplicar un modelo de series de tiempo para predecir el precio del euro-dólar mediante time series forecasting o pronóstico de series de tiempo. Mediante ARIMA que es la mezcla de algunos modelos de series como lo son los modelos autorregresivos de medias móviles e integrados.[2]

Descripción de los datos

El pronóstico multivariante simplemente se diferencia del univariante en que puede hacer el mismo proceso con más de una variable, como se observa a continuación. Donde:

Date, Fecha del dato

Close, es el precio de cierre.

Low, es el precio más bajo en ese día.

High, es el precio más alto en ese día.

Open, es el precio de apertura.

Adj close, precio de cierre ajustado por divisiones y distribuciones de dividendos y / o ganancias de capital.

Volume, Cantidad de un activo concreto en el que se invierte durante un período de tiempo determinado.

Descripción de las fuentes de información

https://alkaline-ml.com/pmdarima/tips_and_tricks.html#period, Este link dio información muy útil y concisa para ajustar los parámetros del modelo.

<https://finance.yahoo.com/quote/EURUSD%3DX/history?p=EURUSD%3DX>, De este link dotó de todos los datos necesarios para entrenar el modelo.

Descripción del preprocesamiento de datos

En la siguiente tabla se expone todas las variables que proporciona Yahoo Finance, dispones de 6 features donde, Volume aparece siempre en cero por lo que no representa ninguna utilidad, Adj Close tiende a tener un valor similar al de close en ocasiones muy particulares varía cuando hay cambio dentro de la compañía que cotiza en bolsa, pero éste nunca es el valor reflejado en la bolsa de valores a la hora de obtener beneficios.

	Open	High	Low	Close	Adj Close	Volume
Date						
2003-12-01	1.203398	1.204007	1.194401	1.196501	1.196501	0.0
2003-12-02	1.196101	1.210903	1.194600	1.208897	1.208897	0.0
2003-12-03	1.209000	1.213003	1.207700	1.212298	1.212298	0.0
2003-12-04	1.212004	1.214403	1.204398	1.208094	1.208094	0.0
2003-12-05	1.207802	1.219096	1.206593	1.218695	1.218695	0.0

Por esas razones esos dos features son eliminados

```
del dataset["Adj Close"]
del dataset["Volume"]
```

El dataset en varias ocasiones presenta valores nulos, esto dificulta a la hora de graficarlo, realizar el entrenamiento y predicción por lo que es importante deshacerse de todas las filas que tengan estos valores nulos

470	2005-09-15,1.227807,1.228607,1.220197,1.223601,1.223601,0
471	2005-09-16,1.223601,1.230603,1.220599,1.223900,1.223900,0
472	2005-09-19,1.215407,1.217597,1.210405,1.213902,1.213902,0
473	2005-09-20,1.213902,1.218606,1.211504,1.211901,1.211901,0
474	2005-09-21,null,null,null,null,null,null
475	2005-09-22,1.216693,1.226904,1.213695,1.215407,1.215407,0
476	2005-09-23,1.215495,1.216205,1.204007,1.204094,1.204094,0
477	2005-09-26,1.202603,1.208007,1.201605,1.207307,1.207307,0
478	2005-09-27,1.207394,1.207904,1.198595,1.201894,1.201894,0

Las siguientes dos líneas de código eliminan por completo las filas defectuosas.

```
dataset=dataset.dropna(how='any',axis=0)#Eliminar filas defectuosas
data_predictc=data_predictc.dropna(how='any',axis=0)
```

Solución del problema

Para pronosticar el precio que tendrá el EUR/USD se implementará el modelo ARIMA o modelo autorregresivo integrado de media móvil, ya que este usa los valores pasados para predecir los del futuro, este consiste en tres componentes los cuales se dividen en tres principalmente:

Autorregresivo (p) -> Número de términos autorregresivos.

Integrado (d) -> Número de diferencias no estacionales necesarias para la estacionariedad.

Media móvil (q) -> Número de errores de pronóstico retrasados en la ecuación de predicción.

Para usar el modelo se utilizará la función `auto_arima` la cual facilita la selección de los parámetros p , d y q . Además, hay un parámetro adicional llamado m el cual se relaciona con el número de observaciones por ciclo estacional, y es uno que debe conocerse a priori. Normalmente, m corresponderá a alguna periodicidad recurrente.[3] En este caso el dataset presenta periodicidad por lo que m será de 12.

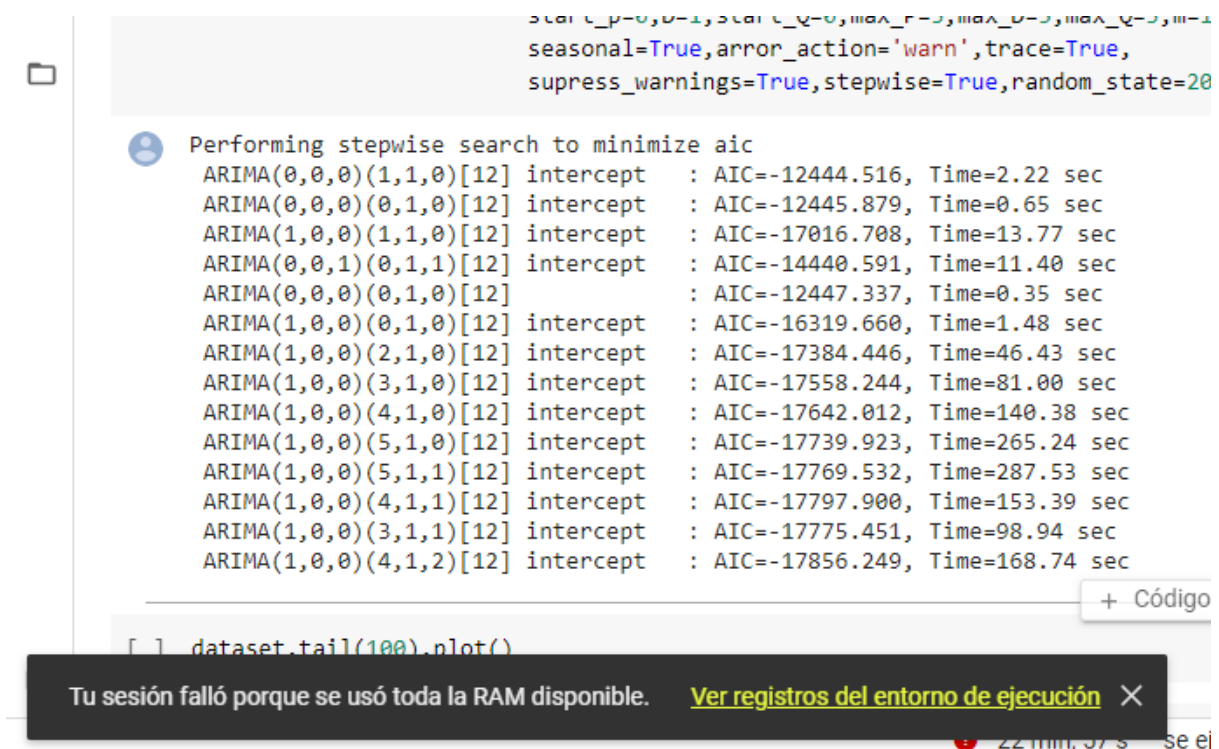
Figura: Modelo ARIMA, tomado de <https://predictivehacks.com/arima-model-in-python/>

En este caso como el EUR/USD tiene periodicidad, la configuración se hará con los parámetros P D Q y m , cómo se va a usar `auto ARIMA` los parámetros P D y Q se configuraran con rango de 0 a 5 y la función encontrará los mejores valores de P , D y Q .

Un parámetro opcional es si se usa el `stepwise` o `Parallelized`, `True` para usar `stepwise` y `False` para usar el modo `parallelized`.

Análisis de resultados

Comenzamos intentando realizar el entrenamiento con todos los datos que teníamos disponibles, pero a la hora de intentar incluir más de 1000 datos, nos encontramos que no contamos con suficiente memoria RAM para completar el entrenamiento. Como se observa en la siguiente imagen.



The screenshot shows a Jupyter Notebook interface. At the top, a code cell contains parameters for `auto_arima`: `start_p=0, start_q=0, max_p=5, max_q=5, m=12, seasonal=True, error_action='warn', trace=True, suppress_warnings=True, stepwise=True, random_state=20`. Below this, a cell displays the output of a stepwise search to minimize AIC, listing various ARIMA models and their corresponding AIC values and execution times. The search concludes with the model `ARIMA(1,0,0)(4,1,2)[12]` having the lowest AIC of -17856.249. At the bottom, a cell contains the command `[] dataset.tail(100).plot()`. A large black error banner at the bottom of the notebook states: "Tu sesión falló porque se usó toda la RAM disponible. Ver registros del entorno de ejecución".

```
start_p=0, start_q=0, max_p=5, max_q=5, m=12, seasonal=True, error_action='warn', trace=True, suppress_warnings=True, stepwise=True, random_state=20
```

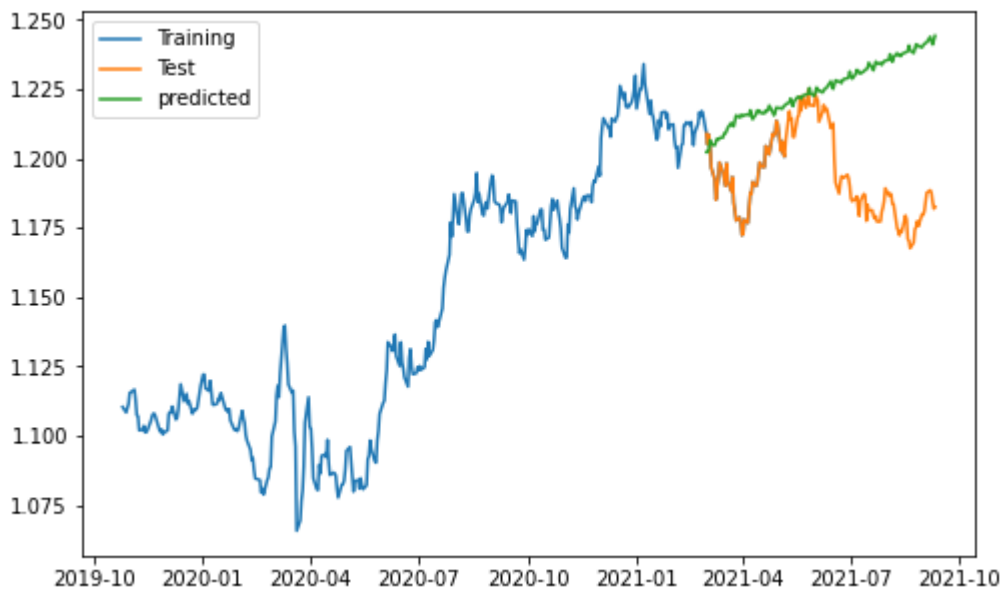
Performing stepwise search to minimize aic

Model	AIC	Time
ARIMA(0,0,0)(1,1,0)[12] intercept	-12444.516	2.22 sec
ARIMA(0,0,0)(0,1,0)[12] intercept	-12445.879	0.65 sec
ARIMA(1,0,0)(1,1,0)[12] intercept	-17016.708	13.77 sec
ARIMA(0,0,1)(0,1,1)[12] intercept	-14440.591	11.40 sec
ARIMA(0,0,0)(0,1,0)[12]	-12447.337	0.35 sec
ARIMA(1,0,0)(0,1,0)[12] intercept	-16319.660	1.48 sec
ARIMA(1,0,0)(2,1,0)[12] intercept	-17384.446	46.43 sec
ARIMA(1,0,0)(3,1,0)[12] intercept	-17558.244	81.00 sec
ARIMA(1,0,0)(4,1,0)[12] intercept	-17642.012	140.38 sec
ARIMA(1,0,0)(5,1,0)[12] intercept	-17739.923	265.24 sec
ARIMA(1,0,0)(5,1,1)[12] intercept	-17769.532	287.53 sec
ARIMA(1,0,0)(4,1,1)[12] intercept	-17797.900	153.39 sec
ARIMA(1,0,0)(3,1,1)[12] intercept	-17775.451	98.94 sec
ARIMA(1,0,0)(4,1,2)[12] intercept	-17856.249	168.74 sec

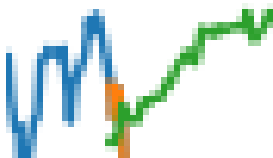
```
[ ] dataset.tail(100).plot()
```

Tu sesión falló porque se usó toda la RAM disponible. Ver registros del entorno de ejecución

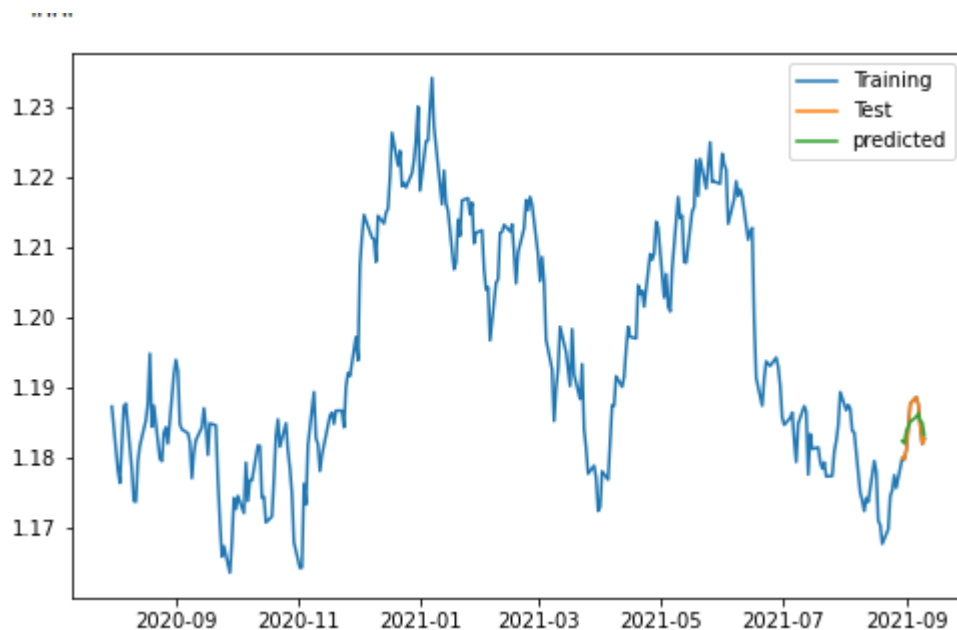
Probamos con 400 datos, donde 140 se utilizan para entrenamiento, por lo que se realizan pronósticos para los siguientes 140 días.



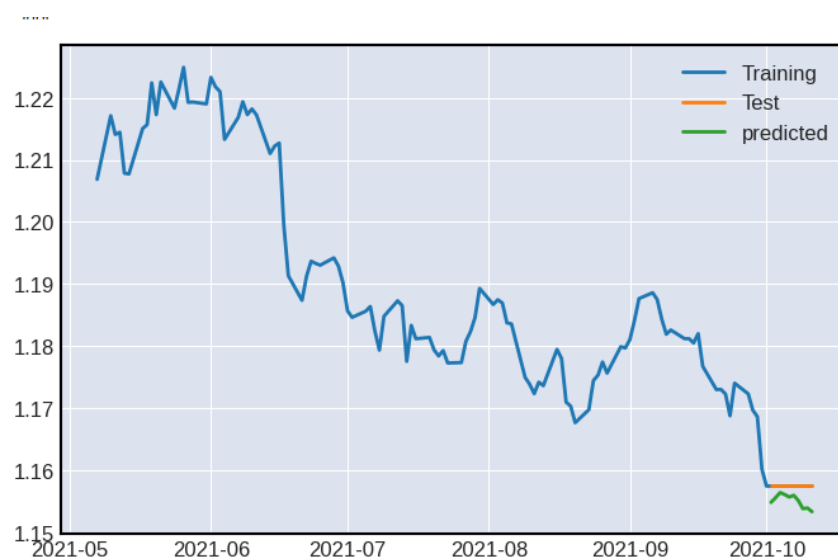
A simple vista parece que la predicción es totalmente errónea, pero, si miramos con más detalle la gráfica, en los primeros 3 o quizás 4 días, se observa como los datos que predice el algoritmo parecen estar muy relacionados con el valor real.



Si probamos con una producción menor, 10 días, observamos como ahora los datos pronosticados son totalmente congruentes.

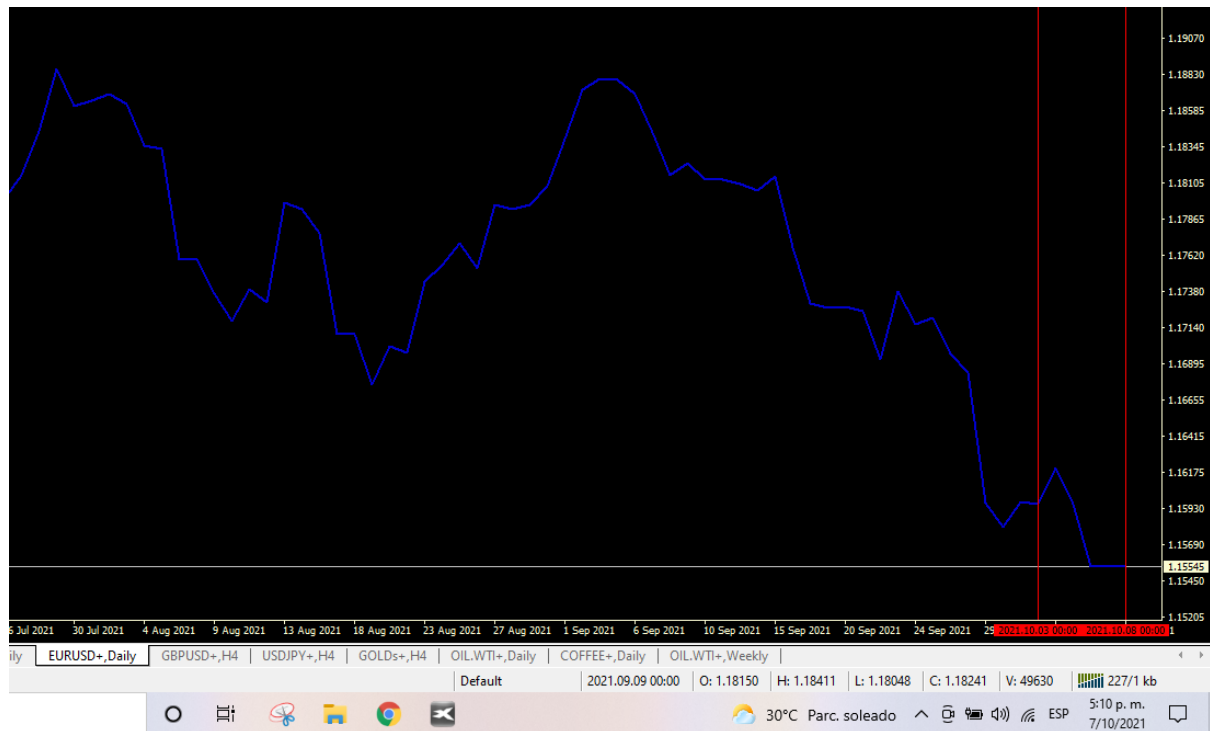


Después de tener resultados satisfactorios con el modelo entrenado, se procede a usar el modelo para realizar realmente una predicción. Nuevamente al entrenarse con los mismos parámetros, pero en esta ocasión, todos los datos más recientes serán para realizar el entrenamiento, y como no se dispone de datos para testing, ya que estos no existen, se dejan de forma ilustrativa el precio del día de hoy constante, (Esto no influye en la predicción) Y se procede a predecir el precio de cierre de los siguiente días, como se observa a continuación.



La predicción como se evidencia en la figura, fue realizada el día 3/10/2021.

La línea verde representa el valor de los precios de cierre , a continuación se muestra los precios que realmente tuvo el eur usd en esos días



Si se comparan las dos gráficas observamos que el pronóstico fue correcto.

El día jueves 7 de octubre se realizó otra predicción, hasta la fecha de hoy solo se puede corroborar el resultado del día 8 y 10 octubre porque el fin de semana el mercado está cerrado. Los resultados obtenidos por el modelo se muestran a continuación.



La línea azul llega hasta el día 7 de octubre y la línea verde es la predicción que realiza el modelo.

La siguiente figura es el comportamiento que realmente está teniendo, aún no se cuenta con suficientes datos para concluir algo sobre esta pronóstico, se deja a los lectores que a la fecha de la lectura corrobore los resultados.



Conclusiones

No es estrictamente necesario que acierte el 100% de las veces ya que esto es prácticamente imposible, pero sí un modelo que permita tener una consistencia a lo largo del tiempo que aunque en ocasiones falle lo importante es que a final de cuentas genere un aumento de los fondos.

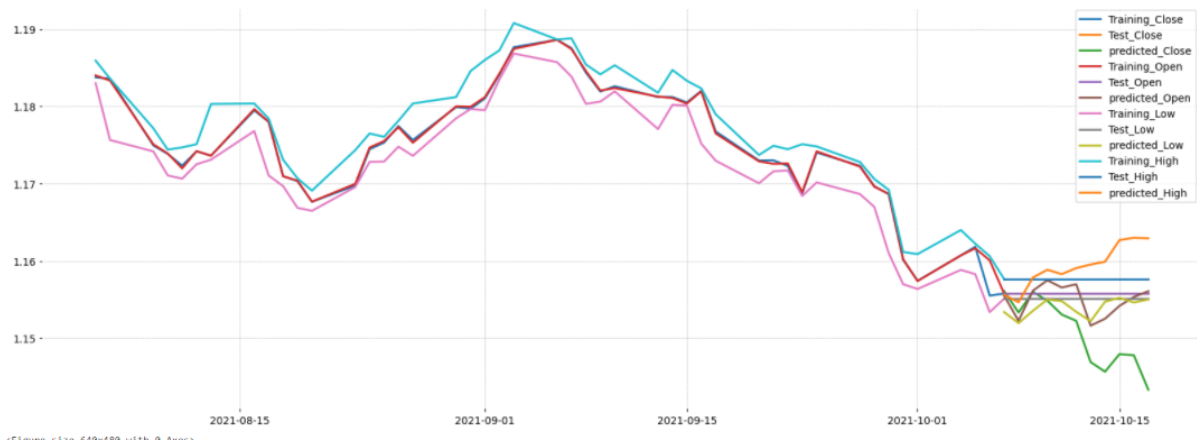
Cuando una tendencia es muy prolongada, provoca que los datos de entrenamiento sean algo similar a una pendiente, produciendo que la predicción sea la continuación de esa recta.

Existe la necesidad de recurrir a más recursos de hardware o un modelo más eficiente, que permita el entrenamiento de un conjunto de datos más grande.

Existen indicadores técnicos que ayudan a analizar mejor el comportamiento de un activo, como lo es el RSI que ayuda a determinar situaciones donde el mercado se encuentra sobrecomprado o sobrevendido, dando como advertencia un cambio de tendencia, dicho indicador, añadido al set de

entrenamiento, podría ayudar a solucionar el problema anteriormente mencionado, sobre tendencias muy prolongadas que provocan una predicción de regresión lineal.

Al intentar realizar una predicción con todas las variables, se produce el siguiente resultado



Al observar la gráfica generada, observamos que las demás variables no tienen un comportamiento coherente, ya que todas las variables siempre tienen que ir acorde con las demás variables. El precio Low, (PredictedL) siempre tiene que ser el valor más bajo, de igual modo precio High (PredictedH) Siempre tiene que ubicarse como el precio más alto. Por lo que se puede concluir que las demás variables no se pueden entrenar con el mismo modelo que los precios de cierre.

REFERENCIAS

[https://www.researchgate.net/publication/28236638_Analisis_de_series_de_tiempo_para_la_prediccion_de_los_precios_de_la_energia_en_la_bolsa_de_Colombia].

[https://www.udape.gob.bo/portales_html/analisisEconomico/analisis/vol14/art02.pdf]

[10]<https://i0.wp.com/datasailors.wpcomstaging.com/wp-content/uploads/2018/12/Sliding-windows-in-time-series-machine-learning-technique.png?zoom=2&resize=353%2C390&ssl=1>

[<https://datasailors.wpcomstaging.com/time-series-as-supervised-learning/>]

[11]Machine Learning for Time Series Forecasting with Python® Francesca Lazzeri, PhD -ISBN: 978-1-119-68236-3

[1]José Francisco López (02 de diciembre, 2018).

Par de divisas. Economipedia.com

[2] <http://www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/363/7/A7.pdf>