

TRADING DE ALTA FRECUENCIA

DATA SCIENCE AND BIG DATA Posgraduate Autor: Juan
CARLOS FORERO CARREÑO

2020 - 2021



UNIVERSITAT_{DE}
BARCELONA



TABLA DE CONTENIDOS

- Introducción y Justificación
- Analisis de datos
- Caso Practico
- Conclusiones y futuras Líneas de Investigación



INTRODUCCIÓN Y JUSTIFICATION

Realizar un montaje exploratorio de series financieras en periodos de muy corto tiempo y utilizar redes neuronales, para predecir el precio de los activos desde una perspectiva de clasificación y regresión.

hacer el back testing de esta estrategia de trading de alta frecuencia un entorno lo más real posible.



CASO PRACTICO

Trading de Alta Frecuencia

Medidas de Riesgo y Desempeño

Algoritmo

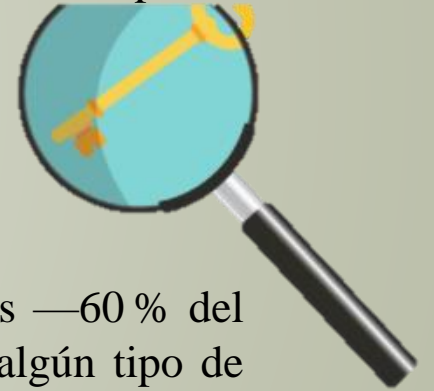
Redes Neuronales Convolucionales

Redes Neuronales LSTM (*long short-term memory networks*)

TRADING DE ALTA FRECUENCIA

Definición

Es un tipo de trading donde las decisiones de inversión tienen un horizonte temporal más pequeño, minutos o incluso segundos, allí las determinaciones de inversión deben ser ejecutadas por algoritmos precisamente por la estrechez de ese horizonte temporal.



Importancia

Se estima que entre el 60 y 70 % de todas las operaciones realizadas en los mercados de acciones —60 % del mercado de futuros y un 50 % en el mercado de treasury bonds americanos— son realizadas por algún tipo de algoritmo. Además, se estima que para el 2026 el valor de las operaciones diarias realizadas por algoritmos alcance US 21.52 billones tomando como base el 2019 con una cifra promedio de US 9.53 billones (Research, 2020).

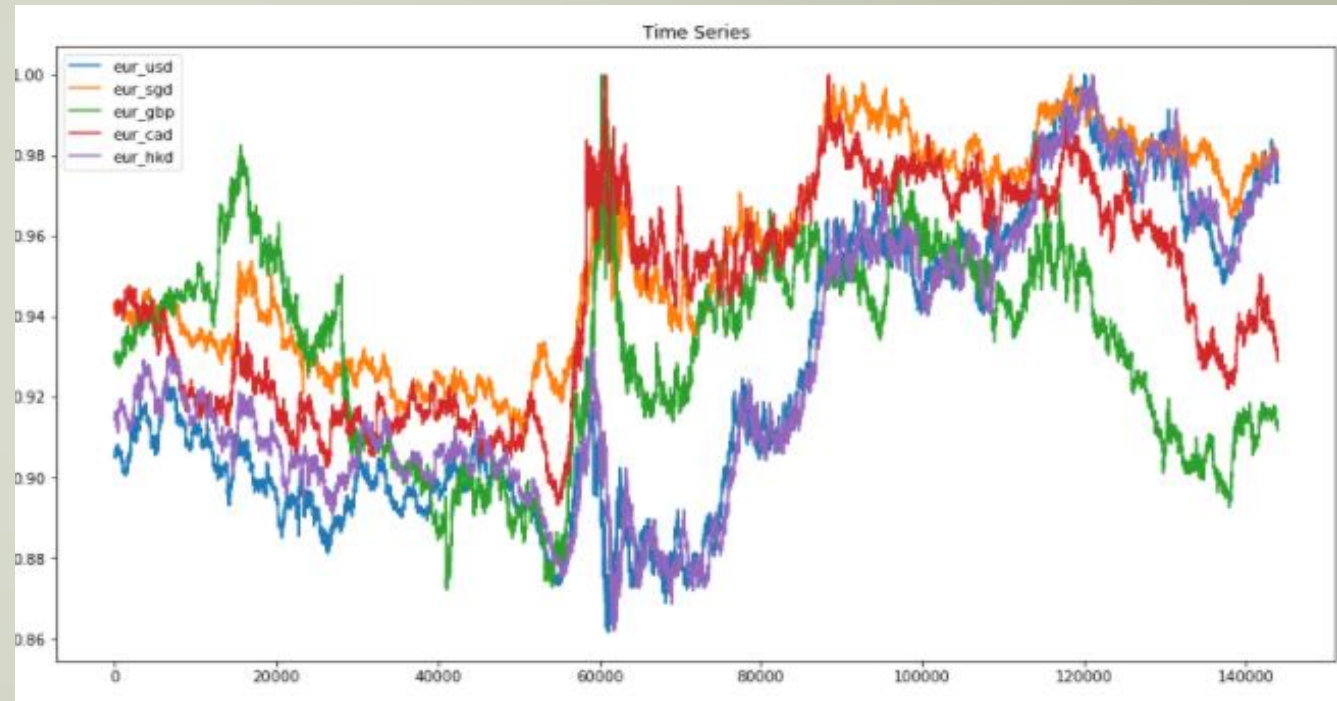
SERIES TEMPORALES Y SELECCIÓN DE PARAMETROS

Fuente de los datos

- Periodo: 23 de mayo del 2019 – 5 de mayo 2021
- Granularidad: 5 minutos
- Total, datos = 140.000 *ticks*
- Datos de validación: 60 % = 84000 *ticks*
- Datos de prueba: 20 % = 28.000
- Datos de validación: 10 % = 14.000

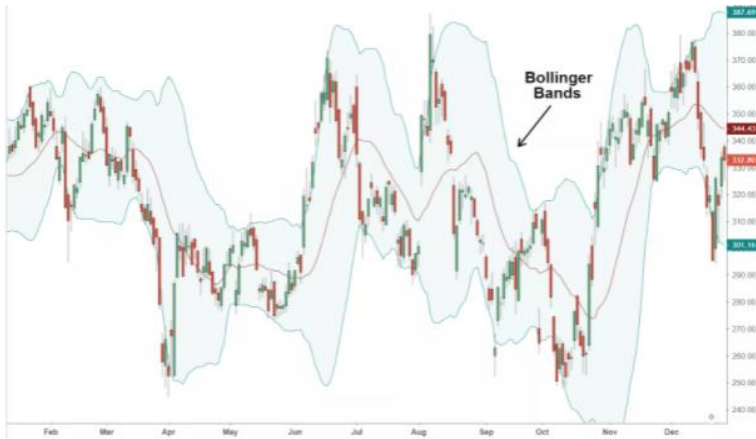
Características principales

- EUR/USD: Euros por dólar americano
- EUR/SGD: Euros por dólar de Singapur
- EUR/GBP: Euros por Libras Esterlinas
- EUR/CAD: Euros por dólar canadiense
- EUR/HKD: Euros por dólar Hongkonés
- Volumen de negociación de cada uno de ellos



CARACTERÍSTICAS ADICIONALES

- Media móvil 10 periodos
- Media móvil 20 periodos
- Media móvil 30 periodos



- RSI: *Relative Strength Index* 14 días
- BBM: Banda de Bollinger media
- BBI: Banda de Bollinger inferior



- BBS: Banda de Bollinger superior
- EMA: Media móvil exponencial 20 días



TRATAMIENTO DE DATOS

Estandarización MinMax

$$x_{rescalado} = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})}$$

Spread Bid – Ask Historico

- 0.6 Puntos básicos = 0.0006

Mínimo retorno requerido

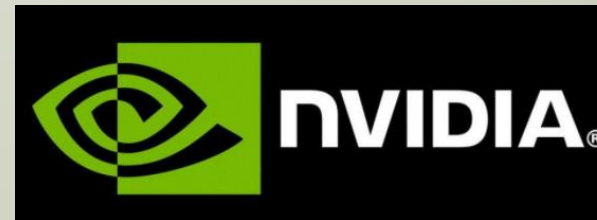
- 0.0006

Ventana móvil

- 15 Periodos hacia atrás
- Numero de características 41
- Numero de prametros $15 \times 41 = 615$

Entrenamiento

- Google Colab Pro
- GPU NVIDIA P100 16GB



MEDIDAS DE RIESGO Y DESEMPEÑO

Retorno total

$$TR_i = \begin{cases} \sum_{i=1}^t c_{i-1} + c_{i-1}(1 + r_i) * k & k = 1 \text{ si tiene posicion en el activo} \\ \sum_{i=1}^t c_{i-1} & k = 0 \text{ si no tiene posicion en el activo} \end{cases}$$

Donde:

TR_i : Retorno total en el momento t

c_{i-1} = valor de la cartera en el momento $t - 1$

r_i : retorno entre el momento t y $t - 1$

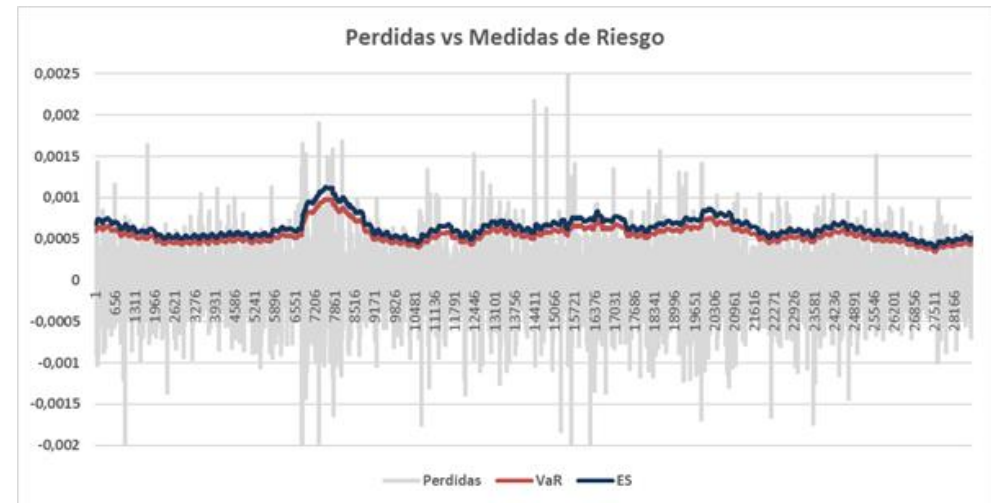
k : define si tenemos el activo en el cartera en el monento t

Value at Risk

$$VaR_{\alpha} = \mu + \sigma \Phi^{-1}(\alpha)$$

Expected shortfall

$$ES_{\alpha} = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha}$$



Medida	Alfa	Perdidas > L	TOTAL	% Error
VaR (L)	99%	281	28785	0,98%
ES (L)	99%	443	28785	1,54%

ALGORITMO

Regresión

$$Compra_i = \begin{cases} compra & r_{i+t} = \ln\left(\frac{f_{i+1}(w)}{x_i}\right) > k \\ no\ hace\ nada\ compra & r_{i+t} = \ln\left(\frac{f_{i+1}(w)}{x_i}\right) \leq k \end{cases} \quad \forall x_i, \forall r_i > 0$$

$$vende_i = \begin{cases} vende & r_i = \ln\left(\frac{f_{i+1}(w)}{x_i}\right) \leq ES \quad \forall x_i, \forall r_i < 0 \\ vende & r_i = \ln\left(\frac{f_{i+1}(w)}{x_i}\right) \leq RT \quad x_i, \forall r_i < 0 \\ vende & r_i = \ln\left(\frac{f_{i+1}(w)}{x_i}\right) < k \quad \forall x_i, \forall r_i < 0 \\ no\ hace\ nada & r_i k < \ln\left(\frac{f_{i+1}(w)}{x_i}\right) \leq 0 \end{cases}$$

• Clasificación

$$Compra_i = \begin{cases} compra & f_{i+1}(w) \in kg_i > k \\ no\ hace\ nada & f_{i+1}(w) \in kg_i < k \end{cases} \quad \forall x_i$$

$$vende_i = \begin{cases} vende & f_{i+1}(w) \in kg_i < RT \quad \forall x_i \forall r_i < 0 \\ vende & f_{i+1}(w) \in kg_i < ES \quad \forall x_i \forall r_i < 0 \\ vende & f_{i+1}(w) \in kg_i < k \quad \forall x_i \forall r_i < 0 \\ no\ hace\ nada & k < f_{i+1}(w) \in kg_i < 0 \end{cases}$$



RESULTADOS

Redes Neuronales LSTM (*long short-term memory networks*)

Red Neuronal Convolucional Completamente conectada con LSTM

Red Neuronal Convolucional Completamente conectada con LSTM Clasificación

Mejor Algoritmo

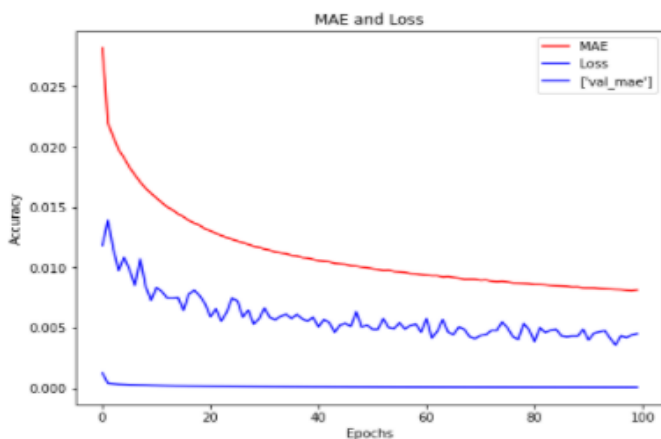
Comparación

RED NEURONALES LSTM (*LONG SHORT-TERM MEMORY NETWORKS*)

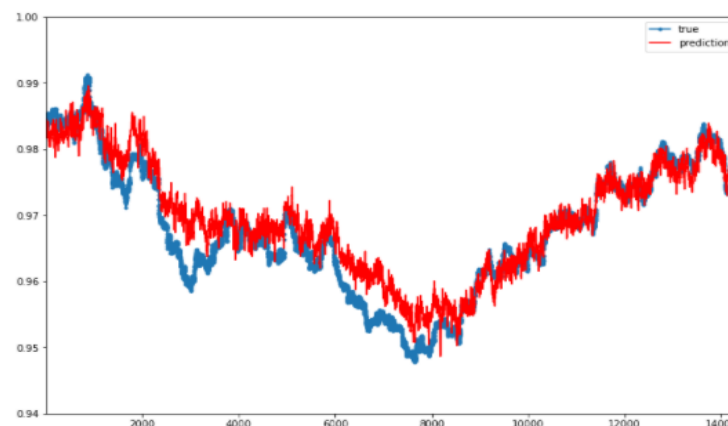
Configuración

1. Capa: LSTM: 900 Neuronas
2. Capa Drop Out : 20%
3. Capa: LSTM: 600 Neuronas
4. Capa Drop Out : 20%
5. Capa: LSTM: 350 Neuronas
6. Capa: Densa

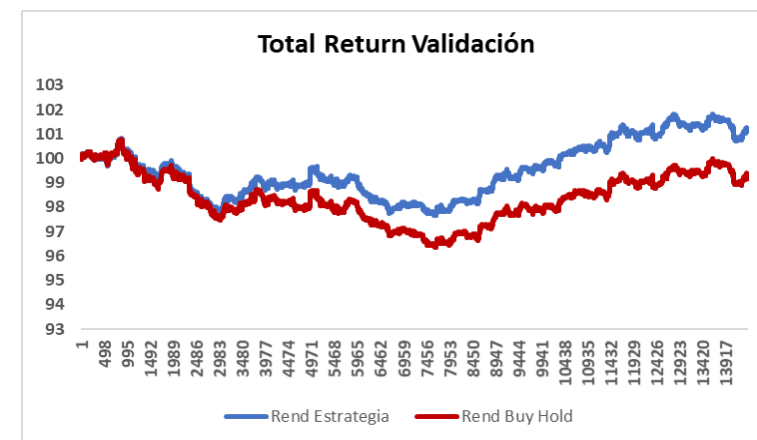
LSTM MAE AND LOSS



Pronostico LSTM validación



Retorno Total



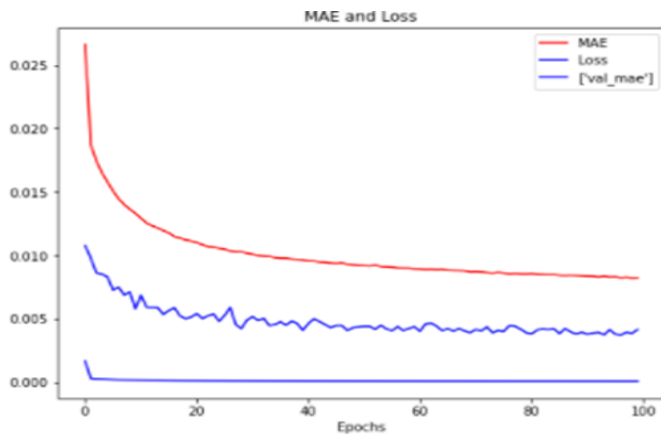
Retorno Anual					
Datos	inicio	Fin	Dias	Estrategia	Buy Hold
Train	23/05/2019	01/10/2020	497,00	no aplica	no aplica
Test	01/10/2020	22/02/2021	144,00	14,77%	8,31%
Validation	22/02/2021	06/05/2021	73,00	5,93%	-3,70%

RED NEURONAL CONVOLUCIONAL COMPLETAMENTE CONECTADA CON LSTM

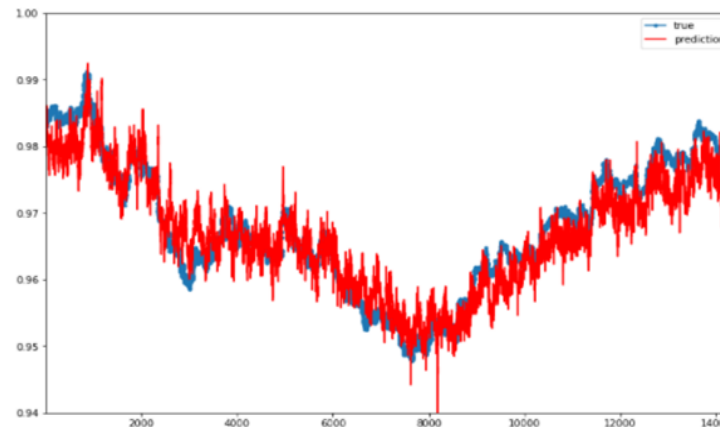
Configuración

1. Capa: Convolucional:Kernel 16*16
2. Capa Drop Out : 20%
3. Capa: LSTM: 650 Neuronas
4. Capa Drop Out : 20%
5. Capa: LSTM: 350 Neuronas
6. Capa: Densa

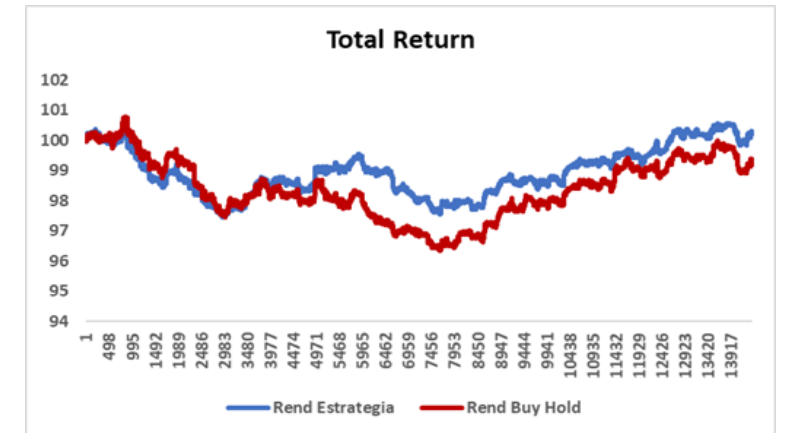
MAE AND LOSS



Pronostico LSTM validación



Retorno Total



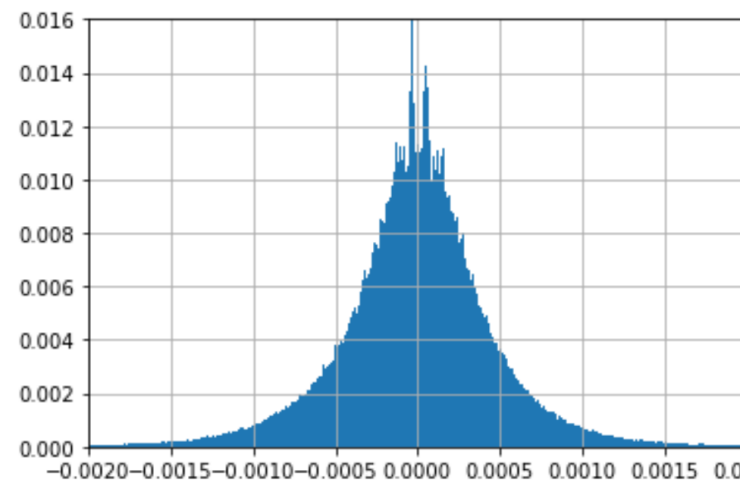
Retorno Anual					
Datos	inicio	Fin	Dias	Estrategia	Buy Hold
Train	23/05/2019	01/10/2020	497,00	no aplica	no aplica
Test	01/10/2020	22/02/2021	144,00	8,86%	8,31%
Validation	22/02/2021	06/05/2021	73,00	1,37%	-3,70%

MODELOS DE CLASIFICACIÓN

Agrupación de retornos en categorías

$$r_i \begin{cases} 0 & \forall, r_i > 0.00006 \\ 1 & -0.00003 \leq r_i \leq 0.00006 \\ 2 & \forall, r_i < -0.00003 \end{cases}$$

r_i : Clasificación del retorno i



Matriz de Confusión

Predicted Values

Positive (1)

Negative (0)

Actual Values

Positive (1)

Negative (0)

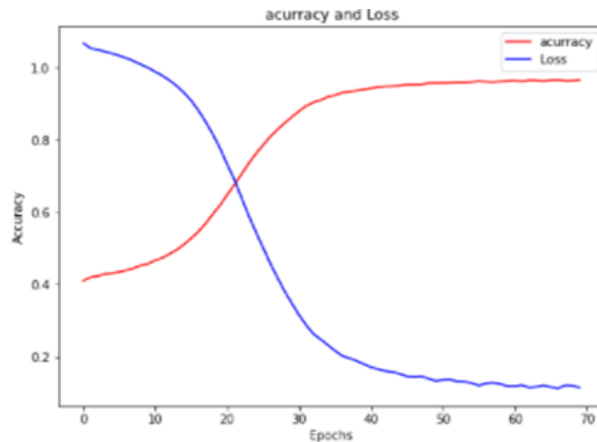
Predicted Values	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

RED NEURONAL CONVOLUCIONAL COMPLETAMENTE CONECTADA CON LSTM (RETORNOS)

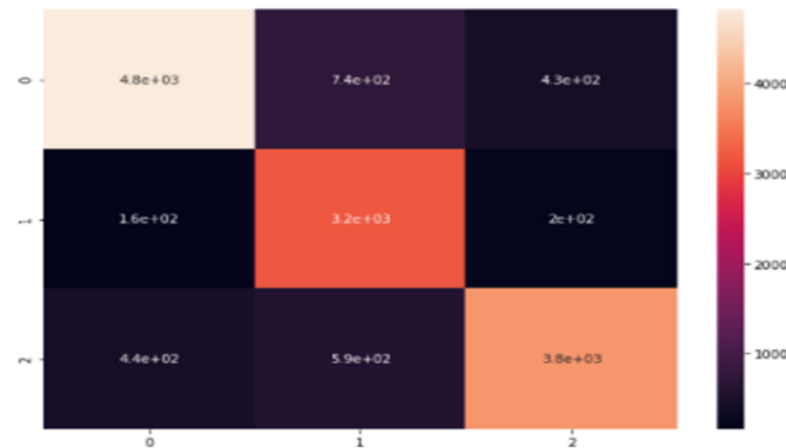
Configuración

1. Capa: Convolucional: Kernel 16*16
2. Capa Drop Out : 20%
3. Capa: LSTM: 650 Neuronas
4. Capa Drop Out : 20%
5. Capa: LSTM: 350 Neuronas
6. Capa: Densa

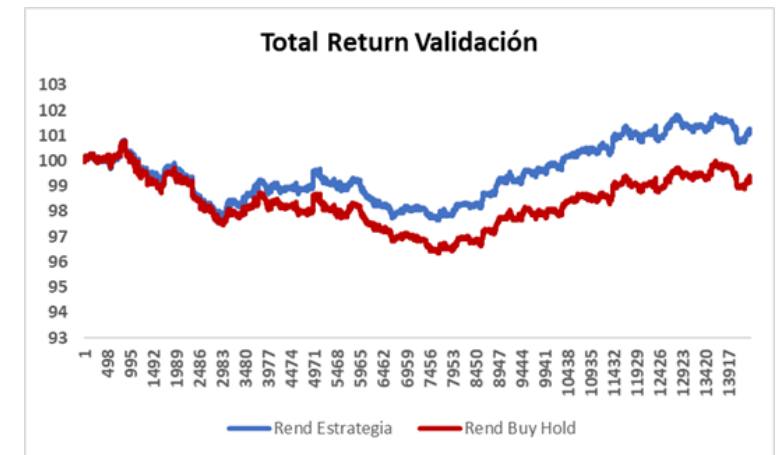
ACURACY AND LOSS



Pronostico validación



Retorno Total



Retorno Anual					
Datos	inicio	Fin	Dias	Estrategia	Buy Hold
Train	23/05/2019	01/10/2020	497,00	no aplica	no aplica
Test	01/10/2020	22/02/2021	144,00	10,79%	8,31%
Validation	22/02/2021	06/05/2021	73,00	5,93%	-3,70%

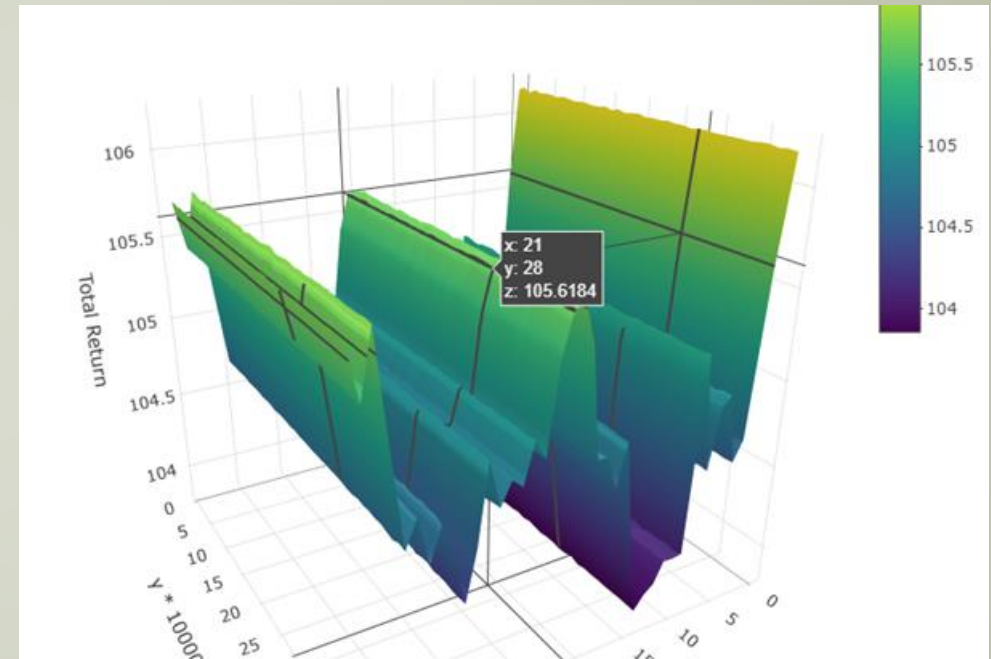
OPTIMIZACIÓN DEL ALGORITMO

Retorno Acumulado venta $RT < 0.009$

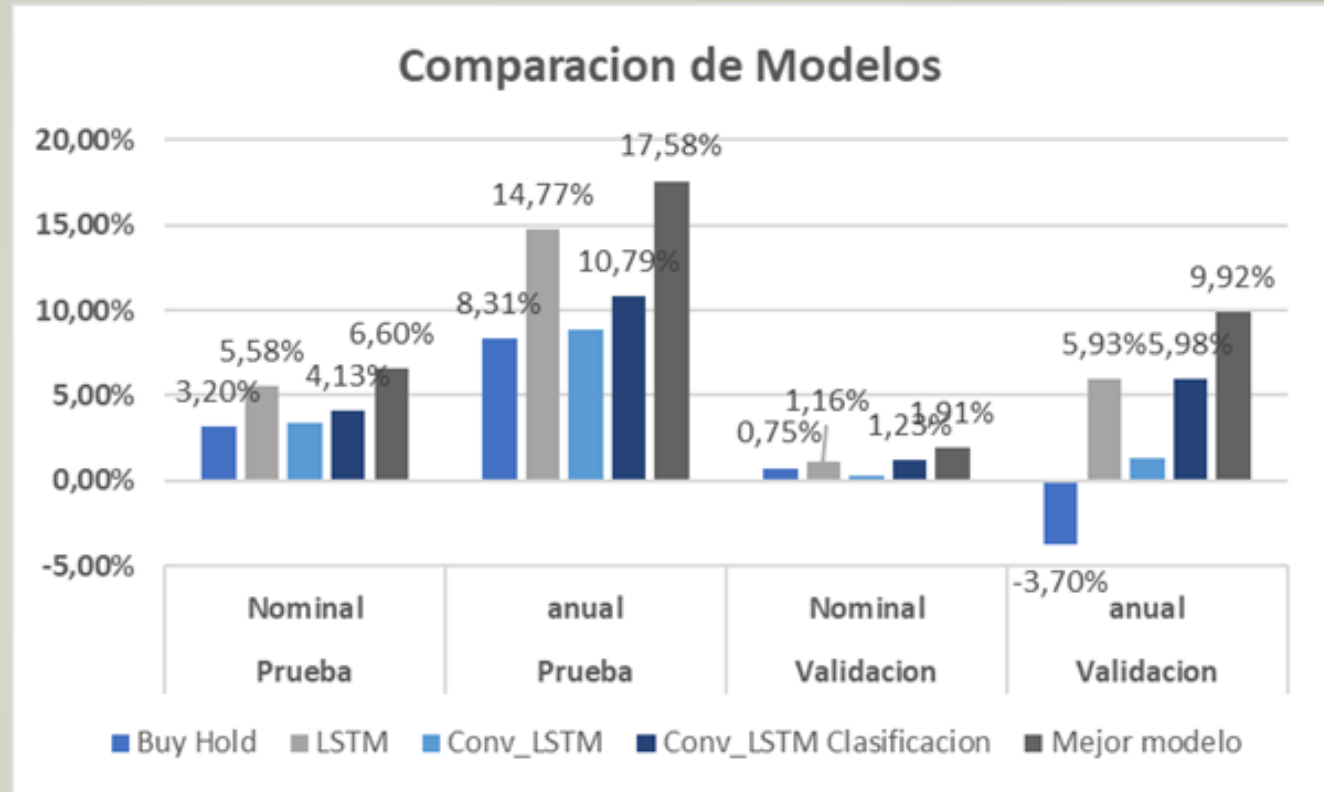
Retorno pronosticado $< ES_i$

Compra a partir de nivel de retorno pronosticado $r_i > 0.0007$

Venta a partir de nivel de retorno pronosticado $r_i < -0.0006$



RESULTADOS



CONCLUSIONES