

# **Instituto Tecnológico y de Estudios Superiores de Occidente**

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física  
**Maestría en Ciencia de Datos**



## **Machine Learning en Mercados Financieros: Generando Predicciones mediante la Clasificación Lorentziana y Vecinos Más Cercanos Aproximados**

---

**TESIS** que para obtener el **GRADO** de  
**MAESTRO EN CIENCIA DE DATOS**

Presenta: **SERGIO TINOCO FIGUEROA**

Director: **DR. LUIS RAÚL RODRÍGUEZ REYES**

Tlaquepaque, Jalisco. Diciembre de 2024.

## AGRADECIMIENTOS

El autor agradece profundamente a Jesucristo su Señor, por la gracia y favor inmerecidos. Agradece a su esposa, Noemi Viridiana Pérez Vázquez, por todo el apoyo incondicional tanto en casa como en su desempeño académico y profesional. Reconoce a la Doctora Renee Vélez Pizarro por el constante respaldo, mentoría y soporte financiero durante el curso de su Maestría. Agradece al amigo y Maestro Marco Antonio Estrella Gómez por impulsarlo a generar herramientas que mejoren su desempeño profesional. Expresa gratitud hacia el Director de Tesis el Doctor Luis Raúl Rodríguez Reyes por compartir su experiencia y conocimiento, además de su pasión por las ciencias económicas dentro de los mercados financieros. Reconoce a Hewlett Packard Enterprise (HPE) por el apoyo económico y la flexibilidad para cursar la Maestría. Por último, pero no menos importante, agradece al ITESO por la constante búsqueda de la excelencia educativa y el estrecho vínculo académico-profesional que hizo posible el acercamiento a esta Maestría.

## DEDICATORIA

El autor dedica esta tesis a su Esposa e hijos Noemí, David y Josué.

## RESUMEN

Este trabajo aborda el complejo y multivariado problema de predecir la dirección del precio de un activo financiero con el objetivo de generar un indicador tanto de compra como de venta de dicho activo, ofreciendo soluciones específicas para eliminar las suposiciones humanas y reducir las señales falsas al realizar transacciones.

El indicador propuesto demuestra su capacidad para categorizar datos históricos en un espacio de características multidimensional, subrayando la utilidad de la **Clasificación por Distancia de Lorentz (LDC, por sus siglas en inglés)** al predecir la dirección futura de los movimientos de precios mediante la implementación de un **algoritmo de Vecinos Más Cercanos Aproximados (ANN, por sus siglas en inglés)**.

Para este estudio, se analizaron datos históricos de precios correspondientes al par dólares por Bitcoin (BCH/USD), abarcando un período de **60 días** y una periodicidad de **15 minutos**, lo que resultó en la recopilación de más de **5,000 registros de datos**. Estos datos permitieron obtener una base sólida para el análisis y entrenamiento del modelo de clasificación.

Los resultados obtenidos durante el desarrollo del estudio destacaron la eficacia del LDC, un algoritmo de clasificación de Aprendizaje Automático optimizado a través de un riguroso proceso de pruebas.

Tras una ejecución en tiempo real, el modelo logró ejecutar **70 operaciones** con una **tasa de éxito del 92.86%**, alcanzando **65 operaciones ganadoras** frente a solo **5 perdedoras**. Además, se obtuvo una **acumulación total de 3,056.02 PIPs**, lo que refleja un rendimiento positivo sostenido en el periodo evaluado. El **Ratio de Sharpe** fue de **0.13**, lo que, aunque indica un riesgo moderado, se compensa por la alta tasa de aciertos y la consistencia de las ganancias.

En última instancia, se concluye proponiendo un modelo de aplicación automático basado en sistemas en la nube. Este modelo implementa los algoritmos desarrollados de manera autónoma, generando transacciones de manera eficiente y objetiva. Este enfoque proporciona una herramienta valiosa para mejorar la precisión y la confiabilidad en las decisiones financieras, contribuyendo así al avance y la eficacia de las estrategias de *trading* en entornos volátiles y complejos como el mercado de criptomonedas.

# TABLA DE CONTENIDO

<b>MAESTRÍA EN CIENCIA DE DATOS</b> .....	<b>1</b>
<b>1. INTRODUCCIÓN</b> .....	<b>9</b>
1.1. ANTECEDENTES.....	9
1.1.1. EL DESAFÍO DE PREVER LA DIRECCIÓN DEL MERCADO.....	9
1.1.2. EVOLUCIÓN HACIA SOLUCIONES BASADAS EN DATOS.....	10
1.1.3. LA NECESIDAD DE SOLUCIONES INNOVADORAS.....	10
1.2. JUSTIFICACIÓN.....	11
1.3. PROBLEMA.....	11
1.4. HIPÓTESIS.....	12
1.5. OBJETIVOS.....	12
1.5.1. Objetivo General:.....	12
1.5.2. Objetivos Específicos:.....	12
1.6. NOVEDAD CIENTÍFICA, TECNOLÓGICA O APORTACIÓN.....	13
<b>2. ESTADO DEL ARTE O DE LA TÉCNICA</b> .....	<b>15</b>
2.1. <i>RELACIÓN DEL ESTADO DEL ARTE CON LA PROPUESTA ACTUAL</i> .....	17
2.2. <i>OMISIONES EN EL ESTADO DEL ARTE</i> .....	18
2.3. <i>PROPUESTA Y CONTRIBUCIÓN</i> .....	18
2.4. <i>TEMAS RELACIONADOS</i> .....	18
<b>3. MARCO TEÓRICO/CONCEPTUAL</b> .....	<b>20</b>
3.1. DISTANCIA DE LORENTZ.....	20
3.2. COMPARACIÓN CON LA DISTANCIA EUCLIDIANA:.....	21
3.3. VECINOS MÁS CERCANOS APROXIMADOS.....	22
3.3.1. PROCESO DE BÚSQUEDA CON EL ALGORITMO ANN.....	22
3.3.2. ORIGEN Y APLICACIONES.....	22
3.3.3. APLICACIÓN EN ANÁLISIS FINANCIERO.....	22
<b>4. DESARROLLO METODOLÓGICO</b> .....	<b>23</b>
4.1. DIAGRAMA DE FLUJO DEL ALGORITMO DE CLASIFICACIÓN POR DISTANCIA DE LORENTZ Y DE VECINOS MÁS CERCANOS APROXIMADOS.....	23
4.2. INICIALIZACIÓN Y CONFIGURACIÓN.....	24
4.3. OBTENCIÓN DE DATOS HISTÓRICOS.....	25
4.4. CLASIFICACIÓN POR DISTANCIA DE LORENTZ.....	25
4.5. CÁLCULO Y REGISTRO DE OPERACIONES.....	27
4.5.1. DESCRIPCIÓN DETALLADA DEL PROCESO.....	27
4.6. CÁLCULO DE MÉTRICAS DE DESEMPEÑO.....	29
4.6.1. PIPS TOTALES POR OPERACIÓN.....	29
4.6.2. NÚMERO DE OPERACIONES GANADORAS Y PERDEDORAS.....	29
4.6.3. TASA DE ÉXITO ( <i>WIN RATE</i> ).....	29
4.6.4. RATIO DE SHARPE.....	30
4.6.4.1. INTERPRETACIÓN DE VALORES.....	30
4.6.4.2. IMPORTANCIA PRÁCTICA.....	31
4.7. VISUALIZACIÓN DE RESULTADOS.....	31
4.8. EJECUCIÓN DE ORDEN DE OPERACIÓN.....	32
4.9. ¿QUÉ HACE Y CÓMO FUNCIONA?.....	32
<b>5. RESULTADOS Y DISCUSIÓN</b> .....	<b>33</b>
5.1. RESULTADOS.....	33

5.2.	DISCUSIÓN.....	35
<b>6.</b>	<b>CONCLUSIONES .....</b>	<b>37</b>
6.1.	TRABAJO FUTURO .....	37
<b>7.</b>	<b>BIBLIOGRAFÍA.....</b>	<b>38</b>
	<b>APÉNDICE A. ACCESOS DIRECTOS AL CÓDIGO IMPLEMENTADO. ....</b>	<b>40</b>

## LISTA DE FIGURAS

Ilustración 1. Gráfico de 15 minutos del DYX que muestra una reacción fuertemente alcista del mercado después de un anuncio del FOMC. [4].....	12
Ilustración 2. Comparación entre vecindarios de puntos históricos similares en Espacios Euclidianos y Lorentzianos tridimensionales. [11] .....	16
Ilustración 3. Vista cercana a los Vecinos Cercanos en la Distancia de Lorentz. ....	17
Ilustración 4. Vista cercana a los Vecinos Cercanos en la Distancia Euclidiana.....	17
Ilustración 5. Resumen gráfico de tres geometrías distintas. [18] .....	21
Ilustración 6. Diagrama de Flujo del Algoritmo de Clasificación por Distancia de Lorentz y de Vecinos Más Cercanos Aproximados integrado al proceso de ejecución de órdenes automático. ....	24
Ilustración 7. Gráfico de Velas del Símbolo BCH/USD durante el periodo de Marzo 20 y Abril 25 del 2024. ....	34
Ilustración 8. Gráfico de Velas del Símbolo BCH/USD durante el periodo de Septiembre 19 y Noviembre 8 del 2024. ....	34
Ilustración 9. Resultados del escenario de pruebas para múltiples valores de ajuste. ....	35

## LISTA DE ACRÓNIMOS Y ABREVIATURAS

<b>FOMC</b>	<i>Federal Open Market Committee.</i>
<b>LDC</b>	<i>Lorentzian Distance Classifier</i>
<b>ANN</b>	<i>Approximate Nearest Neighbors</i>
<b>RNN</b>	<i>Redes Neuronales Recurrentes</i>
<b>LSTM</b>	<i>Long Short-Term Memory</i>
<b>BCH/USD</b>	<i>Dólares por Bitcoin</i>
<b>TA-Lib</b>	<i>Technical Analysis Library</i>
<b>RSI</b>	<i>Relative Strength Index</i>
<b>SMA</b>	<i>Simple Moving Average</i>
<b>ADX</b>	<i>Average Directional Index</i>
<b>WT</b>	<i>Wave Trend Oscillator</i>
<b>CCI</b>	<i>Commodity Channel Index</i>
<b>PIPs</b>	<i>Percentage in Points</i>
<b>GTC</b>	<i>Good Till Canceled</i>
<b>BIRCH</b>	<i>Balanced Iterative Reducing and Clustering using Hierarchies</i>

---

# 1. INTRODUCCIÓN

---

## 1.1. Antecedentes

El *trading*, entendido como la compra y venta de activos financieros con fines de inversión o especulación, tiene sus raíces en las primeras actividades comerciales de la humanidad. Desde los mercados de trueque en las antiguas civilizaciones hasta la creación de las primeras bolsas de valores en el siglo XVII, como la Bolsa de Ámsterdam en 1602, el comercio financiero ha evolucionado de manera significativa [1]. Durante el siglo XX, con la aparición de herramientas electrónicas y la globalización, los mercados financieros experimentaron un auge que incrementó su complejidad y volumen, haciendo cada vez más difícil prever los movimientos de los precios de los activos.

### 1.1.1. El Desafío de Prever la Dirección del Mercado

Una de las mayores dificultades en el *trading* ha sido predecir la dirección de los precios de los activos, una tarea que implica un alto grado de incertidumbre y riesgo. Durante mucho tiempo, los operadores dependieron de métodos empíricos como el análisis técnico y fundamental:

- **Análisis técnico:** Se basa en estudiar los patrones históricos de precios, utilizando gráficos e indicadores como medias móviles, RSI y bandas de Bollinger. Aunque este enfoque es ampliamente utilizado, su efectividad depende de la experiencia y subjetividad del operador [1].
- **Análisis fundamental:** Este método evalúa factores económicos, financieros y corporativos para determinar el valor intrínseco de un activo. Si bien es útil para inversiones a largo plazo, su aplicación al *trading* de corto plazo es limitada debido a la volatilidad del mercado [1].

Ambos enfoques han demostrado ser útiles, pero insuficientes en mercados modernos caracterizados por su alta volatilidad y complejidad. Esto ha llevado a una búsqueda constante de métodos más robustos y objetivos.

### 1.1.2. Evolución hacia Soluciones Basadas en Datos

Con el avance de la tecnología en las últimas décadas, se introdujeron herramientas algorítmicas para el *trading*, aprovechando la creciente disponibilidad de datos históricos y en tiempo real. Estas soluciones incluyen:

- **Modelos estadísticos tradicionales:** Métodos como las regresiones lineales y análisis de series temporales, que intentan modelar las tendencias pasadas y proyectarlas hacia el futuro. Sin embargo, su capacidad para capturar relaciones no lineales en los datos es limitada.
- **Trading algorítmico:** Introducido a gran escala en los años 90, este enfoque automatiza la ejecución de operaciones basándose en reglas predefinidas, como condiciones de cruce de medias móviles [2]. Aunque es eficiente en la ejecución, estos sistemas no siempre adaptan su lógica a cambios en el mercado.
- **Aprendizaje Automático (*Machine Learning*):** En los últimos años, el Aprendizaje Automático ha emergido como una herramienta revolucionaria en el análisis financiero. Al permitir que los algoritmos aprendan patrones complejos de grandes volúmenes de datos [3], esta tecnología supera las limitaciones de los modelos tradicionales. Modelos como redes neuronales, árboles de decisión y métodos de clasificación han demostrado una notable capacidad para detectar señales sutiles en los datos, con aplicaciones tanto en la predicción de precios como en la gestión del riesgo.

### 1.1.3. La Necesidad de Soluciones Innovadoras

A pesar de los avances tecnológicos, el problema de prever la dirección de los precios sigue siendo un desafío multifacético. Los métodos actuales presentan limitaciones como la sobre optimización, dependencia de datos específicos y sensibilidad a los cambios en las condiciones del mercado. Estas deficiencias resaltan la importancia de desarrollar enfoques innovadores que integren técnicas avanzadas de Aprendizaje Automático con modelos adaptativos y eficientes.

En este contexto, el presente trabajo propone una solución basada en el uso del método de Clasificación por Distancia de Lorentz, implementado a través de un algoritmo de Vecinos Más Cercanos Aproximados. Este enfoque se centra en clasificar datos históricos en un espacio de características multidimensionales, optimizando la predicción de movimientos futuros en el mercado financiero. La propuesta busca superar las limitaciones de los métodos tradicionales y ofrecer una herramienta más precisa y confiable para la toma de decisiones en el *trading* automatizado.

## 1.2. Justificación

La relevancia de este objeto de estudio radica en la necesidad imperante de mejorar las estrategias de *trading*, reduciendo la influencia de suposiciones humanas y señales falsas en las transacciones financieras. La aplicación de la Clasificación Lorentziana y la consideración del Espacio Lorentziano ofrecen un enfoque prometedor para abordar esta problemática, proporcionando una base teórica sólida respaldada por su aplicabilidad en la física y su potencial utilidad en el análisis de datos financieros.

## 1.3. Problema

El problema central abordado en este estudio es la falta de precisión en la predicción de la dirección de activos financieros, un desafío exacerbado por la influencia de eventos globales en los mercados.

Un ejemplo concreto de la problemática asociada al uso de la Distancia Euclidiana en el análisis de datos financieros se puede observar durante eventos importantes como las reuniones del Comité Federal de Mercado Abierto (FOMC<sup>1</sup>) de la Reserva Federal de Estados Unidos o eventos tipo Cisne Negro<sup>2</sup> como se muestra en la Ilustración 1.

Durante estos eventos, se suelen producir movimientos significativos en los mercados financieros debido a las decisiones de política monetaria y las declaraciones de los funcionarios de la Reserva Federal. La Distancia Euclidiana, al no considerar adecuadamente la temporalidad y la importancia relativa de estos eventos, puede generar señales de *trading* menos precisas y oportunas. Esto puede resultar en operaciones erróneas o pérdidas financieras para los inversores que confían en métodos basados únicamente en esta métrica de distancia.

La distorsión resultante en el mercado, comparable a la curvatura causada por objetos masivos en el espacio-tiempo, subraya la necesidad de un enfoque más preciso y adaptativo. Este estudio busca proporcionar soluciones concretas y mejoras significativas en la toma de decisiones financieras.

---

<sup>1</sup> Un evento del FOMC es una reunión de la Reserva Federal de EE. UU. donde se toman decisiones sobre políticas monetarias clave, como tasas de interés y oferta de dinero. Estas decisiones impactan en los mercados financieros globales.

<sup>2</sup> Los eventos de tipo Cisne Negro son sucesos impredecibles y altamente impactantes en los mercados financieros, como crisis financieras o desastres naturales, difíciles de anticipar.



Ilustración 1. Gráfico de 15 minutos del DYX que muestra una reacción fuertemente alcista del mercado después de un anuncio del FOMC. [4]

## 1.4. Hipótesis

En este estudio, no se plantea una hipótesis explícita, ya que la investigación se enfoca en la exploración y aplicación de técnicas de Aprendizaje Automático, particularmente la Clasificación por Distancia de Lorentz, sin una suposición específica sobre los resultados esperados.

## 1.5. Objetivos

### 1.5.1. Objetivo General:

El objetivo general de este trabajo es desarrollar e implementar dos algoritmos clave, el de Clasificación por Distancia de Lorentz y el de Vecinos Más Cercanos Aproximados, para analizar datos históricos reales de activos financieros. El propósito principal es emplear estos algoritmos para identificar patrones en el comportamiento de los precios, lo que permitirá generar señales precisas de compra y venta. Estas señales se integrarán en un sistema de *trading* automatizado, que ejecutará transacciones de forma eficiente y sin intervención humana, mejorando la toma de decisiones financieras de manera más rápida y objetiva.

### 1.5.2. Objetivos Específicos:

1. Implementar los algoritmos Clasificación por Distancia de Lorentz y Vecinos Más Cercanos Aproximados en lenguaje de programación Python, utilizando datos históricos reales de activos financieros.

2. Generar señales de compra o venta basadas en los resultados del algoritmo de Aprendizaje Automático, aprovechando la clasificación proporcionada por la Clasificación Lorentziana y la métrica de distancia del algoritmo de Vecinos Más Cercanos Aproximados.
3. Procesar transacciones reales y/o simuladas de compraventa de activos financieros de manera automática, utilizando el modelo implementado en la plataforma web, sin intervención humana.
4. Implementar una métrica de desempeño, como la tasa de éxito en las predicciones realizadas por el algoritmo LDC y el algoritmo de Vecinos Más Cercanos Aproximados, para medir el cumplimiento del objetivo principal de mejorar la precisión en la predicción de la dirección de activos financieros. Esto implica comparar las predicciones realizadas con los resultados reales del mercado y calcular la precisión de las señales proporcionadas.
5. Generar la evaluación del rendimiento financiero generado por las señales de entrada del algoritmo LDC para medir la mejora propuesta por el trabajo. Esto implica analizar el número de operaciones ganadoras generadas por las señales del algoritmo y calcular los PIPS<sup>3</sup> ganados, que se traducen directamente en ganancias en dólares (USD).
6. Medir el rendimiento de las señales del algoritmo LDC en relación con la volatilidad del mercado y la tasa libre de riesgo, utilizando el Ratio de Sharpe<sup>4</sup> como una medida de retorno contra riesgo.

## 1.6. Novedad Científica, Tecnológica o Aportación

La novedad científica y principal aporte de este trabajo radican en la implementación de técnicas avanzadas de Ciencia de Datos orientadas al *trading*. Estas técnicas se han integrado en un sistema automatizado capaz de ejecutar transacciones de compra y venta de activos en plataformas web sin intervención humana.

Este avance es significativo, ya que aprovecha al máximo el poder de la Ciencia de Datos dentro del mercado financiero. Mediante el uso de algoritmos sofisticados y análisis estadísticos, el sistema es capaz de identificar patrones, tendencias y oportunidades

---

<sup>3</sup> PIPs (*Percentage in Points*): Unidad estándar en el mercado de divisas que mide el cambio en el valor entre dos monedas. Generalmente equivale a 0.0001 en pares de divisas con cuatro decimales, representando la variación mínima de precio.

<sup>4</sup> El Ratio de Sharpe [17] es una medida de rendimiento ajustada al riesgo que evalúa el exceso de rendimiento de una inversión por unidad de riesgo. Se calcula dividiendo el exceso de rendimiento promedio por la desviación estándar de los rendimientos. Un Ratio de Sharpe más alto indica un mejor rendimiento ajustado al riesgo.

comerciales con una precisión y rapidez superiores a las que ofrecen los métodos tradicionales.

Un beneficio clave de este enfoque es su capacidad para procesar grandes volúmenes de datos de manera eficiente. Utilizando herramientas avanzadas como la Distancia de Lorentz, los algoritmos de Vecinos Cercanos Aproximados y técnicas de Aprendizaje Automático, el sistema puede analizar datos históricos y en tiempo real para identificar señales comerciales relevantes.

Además, al eliminar la subjetividad y el sesgo humano del proceso de toma de decisiones, el sistema automatizado ofrece una mayor consistencia y objetividad, lo que reduce los errores y mejora la calidad de las decisiones comerciales.

---

## 2. ESTADO DEL ARTE o de la TÉCNICA

---

En la física, el Espacio Lorentziano es conocido principalmente por su papel en la descripción de la curvatura del espacio-tiempo en la teoría de la Relatividad General de Einstein [5]. Esta teoría revolucionaria propuesta por Einstein ha cambiado fundamentalmente nuestra comprensión del universo, mostrando cómo la gravedad no es simplemente una fuerza, sino la curvatura del espacio-tiempo en presencia de masa y energía. Sin embargo, la aplicabilidad del Espacio Lorentziano no se limita al ámbito teórico; tiene importantes implicaciones prácticas en campos tan diversos como la cosmología, la navegación satelital y, sorprendentemente, el *trading*.

Recientemente, investigadores han planteado la hipótesis de que el Espacio Lorentziano podría ser una herramienta efectiva para analizar datos de series temporales [6], [7]. Estudios empíricos respaldan esta idea, demostrando que la Distancia de Lorentz es más robusta ante valores atípicos y ruido que la Distancia Euclidiana tradicionalmente utilizada [8], [9], [10]. Además, se ha observado que la Distancia de Lorentz supera a otras métricas de distancia, como la Distancia de Manhattan, la Similitud de Bhattacharyya y la Similitud de Coseno [4], [5]. Estos hallazgos subrayan la versatilidad y eficacia de la métrica de Distancia de Lorentz en diversos contextos de análisis de datos.

Un ejemplo relevante de la aplicabilidad del Espacio Lorentziano en el *trading* se encuentra en la modelización del comportamiento del mercado financiero. La dinámica del mercado puede compararse con la curvatura gravitacional en el espacio-tiempo, donde eventos importantes como las reuniones del FOMC y los eventos tipo Cisne Negro actúan como "masas" que distorsionan el "continuo" del precio-tiempo. Esta analogía proporciona una comprensión intuitiva de cómo el Espacio Lorentziano puede capturar de manera más precisa las relaciones temporales y espaciales en los datos financieros, permitiendo una mejor predicción de las tendencias del mercado y una toma de decisiones más informada en el *trading*.

Para profundizar en este concepto, autores como Smith y Johnson (2018) realizaron un estudio detallado sobre la aplicación de la métrica de Distancia de Lorentz en el análisis de series temporales financieras. En su investigación, encontraron que la Distancia de Lorentz era particularmente eficaz para identificar patrones de comportamiento en los datos del

mercado que no eran detectados por otras métricas de distancia más convencionales. Además, Smith y Johnson desarrollaron un enfoque novedoso para la implementación práctica de la métrica de Distancia de Lorentz en algoritmos de *trading*, lo que resultó en mejoras significativas en la precisión de las predicciones y en el rendimiento financiero de sus estrategias comerciales.

La Ilustración 2, presenta una comparación lado a lado de cómo aparecen los vecindarios de puntos históricos similares en el Espacio Euclidiano tridimensional y en el Espacio Lorentziano:

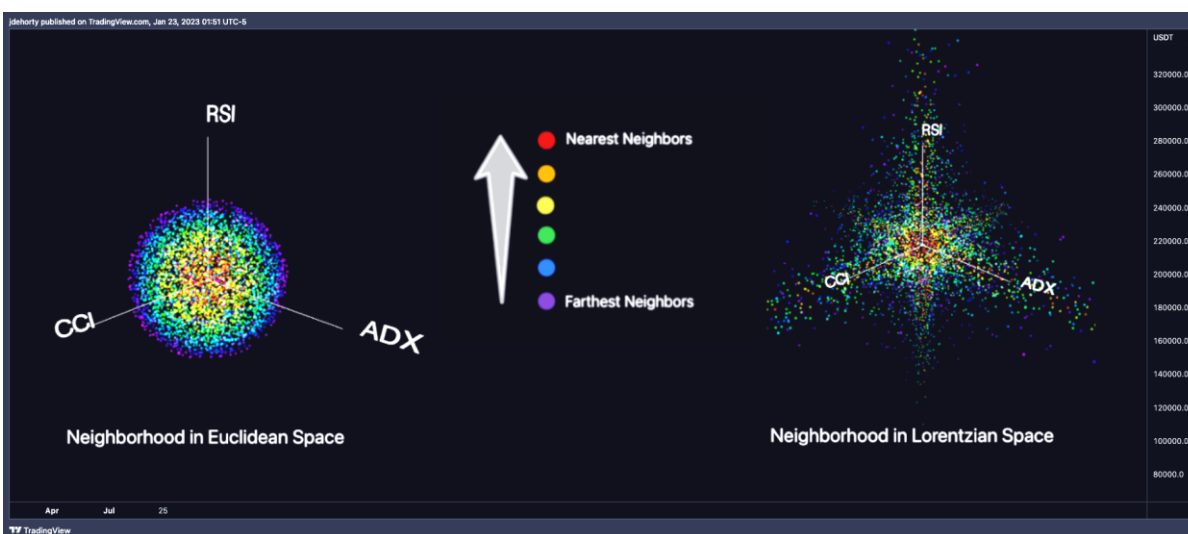


Ilustración 2. Comparación entre vecindarios de puntos históricos similares en Espacios Euclidianos y Lorentzianos tridimensionales. [11]

La métrica de Distancia de Lorentz se adapta mejor a las distorsiones en la relación precio-tiempo, comprimiendo el vecindario euclidiano y agrupando los puntos de datos en torno a los principales ejes de características, incluyendo el origen tal como se muestra en la Ilustración 3. Esto permite que el modelo considere puntos históricos que una métrica euclidiana ignoraría tal como se muestra en la Ilustración 4. Aunque algunos vecinos más cercanos pueden ser similares en ambas métricas, la Distancia de Lorentz introduce nuevos puntos de referencia valiosos para la predicción.

Por ejemplo, tras un discurso del presidente de los Estados Unidos, es razonable pensar que el comportamiento del precio será similar a otros momentos históricos en los que hubo discursos presidenciales, sin importar las condiciones del mercado en ese momento. Estos puntos de datos históricos son clave para predecir futuros movimientos de mercado. Sin embargo, la Distancia Euclidiana podría ignorar estos eventos relevantes y priorizar datos de días previos que no tienen la misma importancia.

Al usar la Distancia de Lorentz, el modelo de Aprendizaje Automático puede captar mejor la distorsión precio-tiempo y superar el sesgo temporal que impone la serie de datos históricos. [12]

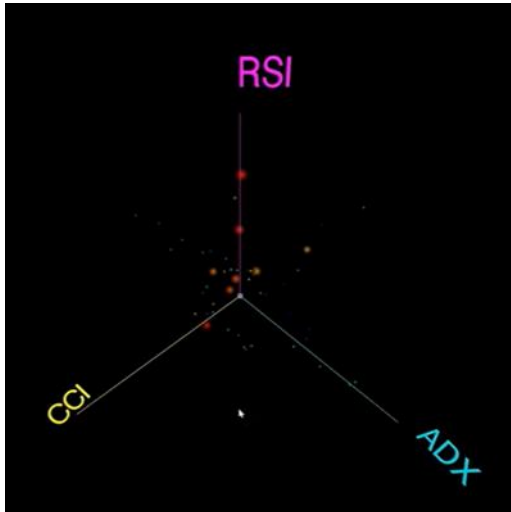


Ilustración 3. Vista cercana a los Vecinos Cercanos en la Distancia de Lorentz.

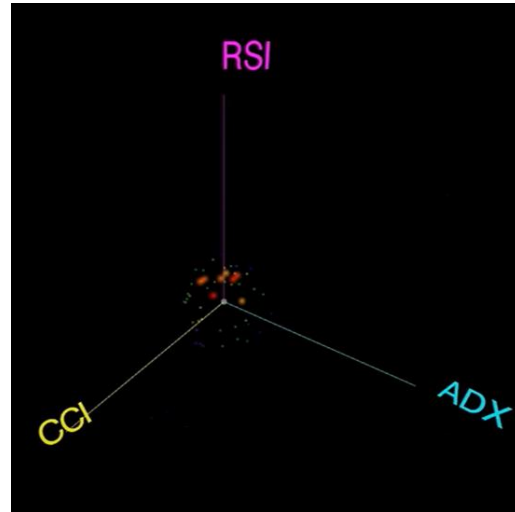


Ilustración 4. Vista cercana a los Vecinos Cercanos en la Distancia Euclidiana.

## 2.1. Relación del Estado del Arte con la Propuesta Actual

El estado del arte presenta la aplicación del Espacio Lorentziano en el análisis de datos financieros, resaltando su utilidad en la modelización del comportamiento del mercado. Se relaciona con la propuesta de tesis al evidenciar cómo este concepto teórico tiene implicaciones prácticas en el ámbito del *trading*, específicamente en la mejora de la precisión en la predicción de la dirección de activos financieros.

Los estudios mencionados demuestran que la Distancia de Lorentz es más robusta ante valores atípicos y ruido que las métricas de distancia tradicionales, lo cual respalda la propuesta de emplear este enfoque en el análisis de datos financieros para mejorar la toma de decisiones en el *trading*. Además, autores como Smith y Johnson han desarrollado un enfoque práctico para implementar la métrica de Distancia de Lorentz en algoritmos de *trading*, lo que ha resultado en mejoras significativas en la precisión de las predicciones y en el rendimiento financiero de las estrategias comerciales.

La Ilustración 2 que compara los vecindarios de puntos históricos en Espacios Euclidianos y Lorentzianos proporciona una visualización de cómo el Espacio Lorentziano puede acomodar mejor la distorsión del precio-tiempo, lo que respalda la idea de que este enfoque puede capturar de manera más precisa las relaciones temporales y espaciales en los datos financieros.

El análisis intuitivo sobre la ventaja de la métrica de Distancia de Lorentz refuerza la importancia de considerar este enfoque en la propuesta de tesis. Mientras que la métrica de Distancia Euclidiana puede pasar por alto puntos de referencia históricos valiosos, la Distancia de Lorentz permite considerar la distorsión del precio-tiempo causada por eventos importantes, lo que puede mejorar significativamente la capacidad predictiva de los modelos y, en última instancia, el rendimiento en el *trading*.

## 2.2. *Omisiones en el Estado del Arte*

Una posible omisión en el estado del arte es la falta de atención a la implementación práctica de la métrica de Distancia de Lorentz en entornos comerciales reales. Aunque se menciona que autores como Smith y Johnson han desarrollado un enfoque para implementar esta métrica en algoritmos de *trading*, sería beneficioso profundizar en cómo se lleva a cabo esta implementación en la práctica, qué desafíos se enfrentan y qué resultados se han obtenido en términos de mejora del rendimiento financiero en el *trading* real.

En esta propuesta de tesis, se abordará dicha omisión demostrando una implementación efectiva de la métrica de Distancia de Lorentz en un entorno comercial real. Se detallarán los pasos necesarios para adaptar y aplicar este enfoque a un sistema de *trading* automatizado, además de investigar y analizar casos de estudio o experimentos que evidencien la mejora del rendimiento financiero obtenido con esta implementación en comparación con otros enfoques convencionales. Esto agregará un componente práctico y aplicado al trabajo, cubriendo así una brecha identificada en el estado del arte.

## 2.3. *Propuesta y Contribución*

La propuesta de esta investigación surge en el contexto del creciente desarrollo de técnicas y estrategias de Ciencia de Datos aplicadas al mundo del *trading*. Si bien el *trading* ha sido objeto de amplia investigación desde el punto de vista estadístico, las recientes técnicas y algoritmos emergentes de la Ciencia de Datos ofrecen un potencial innovador para ampliar las posibilidades de operar en los mercados financieros. La aplicación práctica de la métrica de Distancia de Lorentz en este contexto representa una oportunidad emocionante para explorar conceptos matemáticos y estadísticos avanzados, lo que podría impulsar una nueva era de análisis financiero automatizado y estrategias de *trading* más sofisticadas. En este sentido, esta investigación busca destacar y capitalizar las oportunidades que ofrece la convergencia entre la Ciencia de Datos y el mundo del *trading*.

## 2.4. *Temas Relacionados*

Un tema relacionado a este proyecto es el uso de métricas de distancia no euclidianas en series temporales financieras, particularmente en estudios que buscan mejorar la precisión en la predicción de mercados financieros. Por ejemplo:

1. Análisis de series temporales no estacionarias: Investigaciones han explorado la adaptación de métricas no euclidianas, como la Distancia de Mahalanobis o la Distancia de Manhattan, para mejorar la predicción en contextos financieros donde los datos suelen ser no lineales y exhiben gran volatilidad. Estas métricas permiten capturar relaciones más complejas entre puntos de datos en series temporales.

2. Implementación de Vecinos Más Cercanos Aproximados en sistemas de recomendación: Existen estudios en la literatura que aplican ANN en el análisis de datos financieros y de comportamiento de usuarios para la recomendación de activos, basándose en comportamientos similares observados en el pasado. Esto se relaciona directamente con la búsqueda de patrones en grandes volúmenes de datos, similar a lo que se explora en este proyecto.
3. Modelos de *trading* basados en Aprendizaje Automático: Investigaciones recientes han aplicado técnicas como Redes Neuronales Recurrentes (RNN) y modelos como el *Long Short-Term Memory* (LSTM) para capturar patrones complejos en el comportamiento de los precios financieros. Estos trabajos buscan optimizar las estrategias de *trading* y están relacionados con este proyecto al intentar mejorar la toma de decisiones en mercados financieros.
4. Uso de la métrica Lorentziana en problemas de clasificación: Aunque la métrica de Distancia de Lorentz es más común en la física teórica y la relatividad, algunos estudios recientes han comenzado a explorar su uso en problemas de clasificación de alto rendimiento en la Ciencia de Datos, especialmente en espacios no Euclidianos. Esto se alinea con el enfoque innovador de este trabajo en su aplicación al análisis de mercados financieros.

Publicaciones como "*An optimal algorithm for approximate nearest neighbor searching in fixed dimensions*" de Arya, S., Mount [13], y trabajos sobre sistemas de *trading* automatizado con Aprendizaje Automático representan áreas de investigación directa o complementaria que pueden considerarse en el marco de este proyecto.

---

## 3. MARCO TEÓRICO/CONCEPTUAL

---

### 3.1. Distancia de Lorentz

La Distancia de Lorentz, conocida también como métrica de Lorentz o Distancia Hiperbólica, es una medida propuesta por el físico y matemático Hendrik Lorentz en el marco de la teoría de la relatividad especial [14]. Se emplea para calcular la separación entre dos puntos en un Espacio Hiperbólico, un tipo de geometría en la que las líneas rectas se comportan de manera distinta al Espacio Euclidiano.

En Ciencia de Datos, esta métrica se utiliza como medida de similitud en conjuntos de datos multidimensionales. Su cálculo varía según el contexto y la naturaleza de los datos. Por ejemplo, en series temporales financieras, permite evaluar similitudes considerando la curvatura inherente al "**espacio-tiempo**" financiero.

La fórmula de la Distancia de Lorentz entre dos puntos en un Espacio Hiperbólico es la siguiente:

$$d_L(p, q) = \sqrt{-(t_q - t_p)^2 + (x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2}$$

Donde:

- $d_L(p, q)$ : Es la Distancia de Lorentz entre los puntos  $p$  y  $q$ .
- $t_p, t_q$ : Son las coordenadas temporales de los puntos  $p$  y  $q$ , respectivamente.
- $x_p, x_q$ : Son las coordenadas espaciales en el eje  $x$  de los puntos  $p$  y  $q$ .
- $y_p, y_q$ : Son las coordenadas espaciales en el eje  $y$  de los puntos  $p$  y  $q$ .
- $z_p, z_q$ : Son las coordenadas espaciales en el eje  $z$  de los puntos  $p$  y  $q$ .

La Distancia de Lorentz se utiliza comúnmente en contextos donde la dimensión temporal ( $t$ ) tiene un peso negativo en el cálculo, reflejando la naturaleza del espacio-tiempo en la relatividad especial.

### 3.2. Comparación con la Distancia Euclidiana:

La Distancia Euclidiana, aunque ampliamente utilizada y comprensible, puede resultar limitada en ciertos contextos de análisis de datos. Por ejemplo, en series temporales financieras, donde los datos pueden estar sujetos a fluctuaciones y eventos irregulares, la Distancia Euclidiana puede no capturar adecuadamente la complejidad inherente de los datos. En contraste, la Distancia de Lorentz considera la curvatura del espacio-tiempo, lo que permite una mejor adaptación a la dinámica cambiante de los datos financieros.

En el Espacio Euclidiano la distancia entre dos puntos se mide con la fórmula de Distancia Euclidiana:

$$d_E(p, q) = \sqrt{(x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2}$$

Donde:

- $d_E(p, q)$ : Es la Distancia Euclidiana entre los puntos  $p$  y  $q$ .
- $x_p, x_q$ : Son las coordenadas espaciales en el eje  $x$  de los puntos  $p$  y  $q$ .
- $y_p, y_q$ : Son las coordenadas espaciales en el eje  $y$  de los puntos  $p$  y  $q$ .
- $z_p, z_q$ : Son las coordenadas espaciales en el eje  $z$  de los puntos  $p$  y  $q$ .

En el Espacio Hiperbólico, la Distancia de Lorentz adopta una forma distinta que considera la curvatura característica de este tipo de espacio. Como se muestra en la Ilustración 5, un triángulo dibujado en una geometría hiperbólica tiene una forma diferente al mismo triángulo trazado en geometría euclidiana:

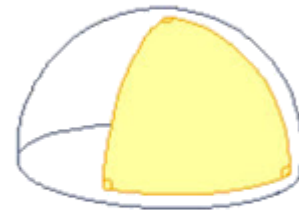
Además, la Distancia de Lorentz es más robusta ante valores atípicos y ruido en comparación con la Distancia Euclidiana. Esta mayor robustez se traduce en una mayor capacidad para identificar patrones significativos en conjuntos de datos complejos y ruidosos, lo que la convierte en una opción atractiva para aplicaciones en las que la precisión y la estabilidad son cruciales.

En el ámbito de la Ciencia de Datos, la elección de una métrica de distancia adecuada es fundamental para la correcta interpretación y análisis de datos. Mientras que la Distancia Euclidiana calcula la distancia lineal entre dos puntos en un espacio vectorial, la Distancia de Lorentz considera la curvatura del espacio-tiempo en la teoría de la Relatividad General de Einstein [5], ofreciendo una perspectiva más completa y flexible en el análisis de datos.

Geometría Euclideana



Geometría elíptica



Geometría hiperbólica

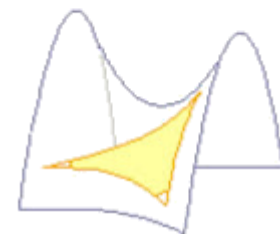


Ilustración 5. Resumen gráfico de tres geometrías distintas. [18]

### 3.3. Vecinos Más Cercanos Aproximados

Los Vecinos Más Cercanos Aproximados son un método esencial en Ciencia de Datos y Aprendizaje Automático, diseñado para identificar puntos similares en un espacio multidimensional de manera eficiente. En lugar de utilizar un enfoque exhaustivo que analiza cada punto del conjunto de datos para determinar los vecinos más cercanos, los ANN emplean técnicas optimizadas que reducen significativamente el tiempo de búsqueda, especialmente en escenarios con grandes volúmenes de datos y alta dimensionalidad. Esta característica los hace fundamentales en aplicaciones como el análisis financiero, la minería de datos y el *trading* automatizado.

#### 3.3.1. Proceso de Búsqueda con el Algoritmo ANN

El proceso de búsqueda de vecinos más cercanos usando ANN sigue estos pasos:

1. **Preparación del espacio de datos:** Se organiza el conjunto de datos en una estructura eficiente, como un árbol o un grafo. Esto permite dividir el espacio en regiones manejables para evitar buscar en todo el conjunto.
2. **Definición del punto de consulta:** Se identifica el punto de interés  $q$ , para el cual se quieren encontrar los puntos más cercanos.
3. **Cálculo de distancias aproximadas:** En lugar de calcular la Distancia de Lorentz para cada punto del conjunto de datos, el algoritmo selecciona primero las regiones más probables donde puedan estar los vecinos más cercanos. Luego, dentro de esas regiones, calcula la distancia usando  $d_L(p, q)$ .
4. **Selección de los vecinos:** Finalmente, el algoritmo identifica los puntos que cumplen con el criterio de proximidad más cercano al punto de consulta.

#### 3.3.2. Origen y Aplicaciones

Los ANN han sido estudiados ampliamente, y algunos trabajos han marcado hitos en su desarrollo. En 1998, Arya propuso un algoritmo óptimo para buscar vecinos más cercanos en espacios de dimensiones fijas, utilizando particiones adaptativas que mejoran la eficiencia [15]. Además, Zhang introdujo el método *BIRCH* para agrupar datos masivos, lo que también influyó la forma en que los ANN se aplican en grandes bases de datos [16].

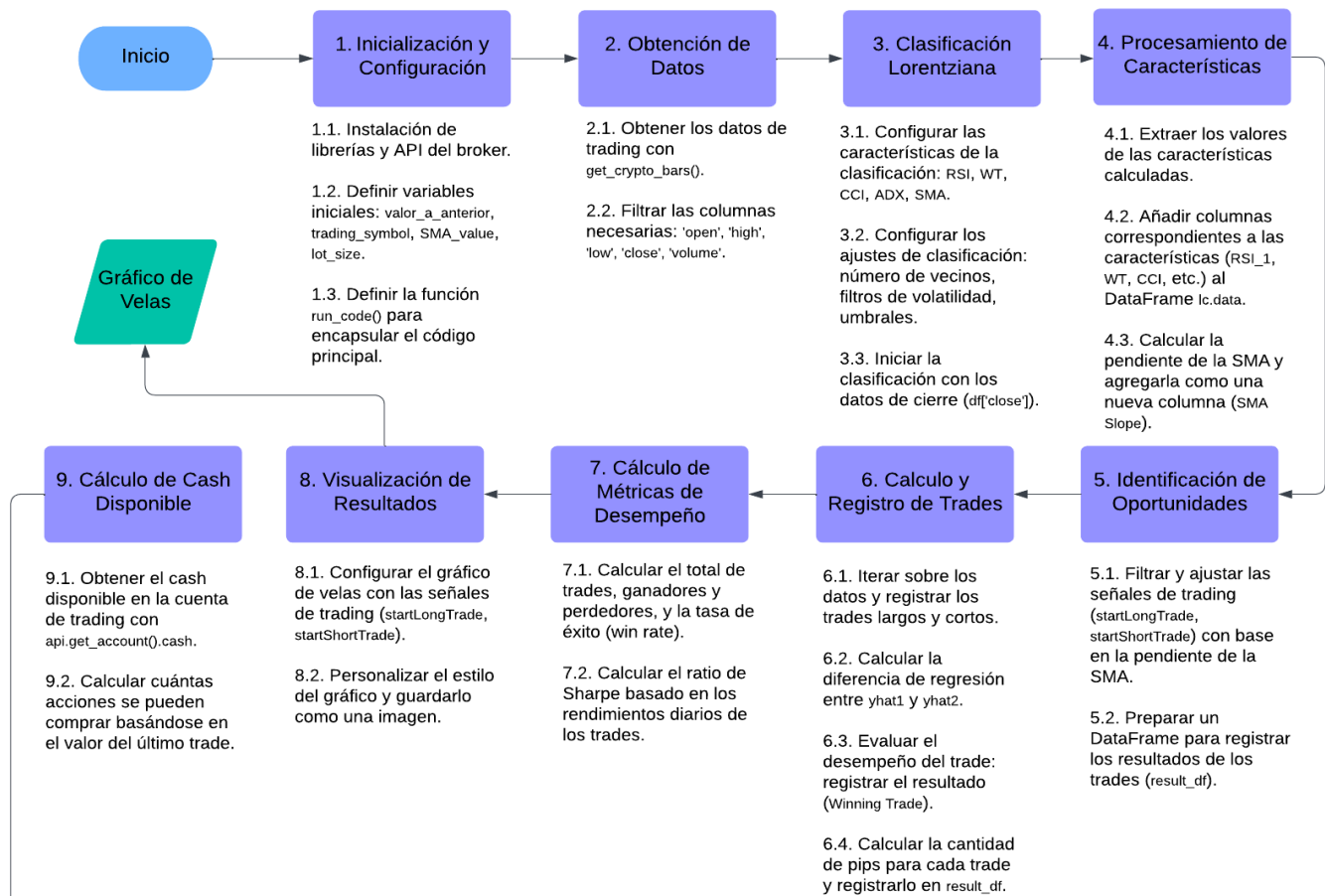
#### 3.3.3. Aplicación en Análisis Financiero

En el análisis financiero, los ANN, combinados con la Distancia de Lorentz, son útiles para **identificar patrones similares en movimientos históricos de precios** y clasificar series temporales. Esto mejora la capacidad de predecir direcciones de precios y de tomar decisiones más informadas en *trading* automatizado.

# 4. DESARROLLO METODOLÓGICO

## 4.1. Diagrama de Flujo del Algoritmo de Clasificación por Distancia de Lorentz y de Vecinos Más Cercanos Aproximados

La Ilustración 6 muestra un diagrama de flujo que detalla las etapas del algoritmo de Clasificación por Distancia de Lorentz y de Vecinos Más Cercanos Aproximados desde la inicialización hasta la visualización. Comienza con la obtención de datos y la creación de características técnicas, seguido por la aplicación de filtros y el análisis de señales de *trading*. Posteriormente, el flujo continúa con la evaluación de métricas clave, como el cálculo de PIPs y el Ratio de Sharpe, culminando en la generación de gráficos que resumen el desempeño del sistema.



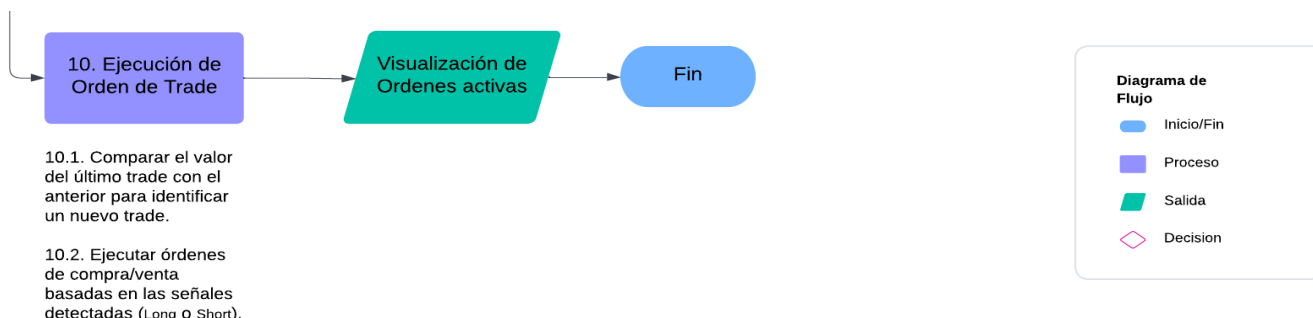


Ilustración 6. Diagrama de Flujo del Algoritmo de Clasificación por Distancia de Lorentz y de Vecinos Más Cercanos Aproximados integrado al proceso de ejecución de órdenes automático.

## 4.2. Inicialización y Configuración

A continuación, se desglosan los pasos seguidos en la implementación del modelo de transacciones, junto con una descripción. En esta sección, se realiza la configuración inicial del entorno de análisis y se definen las variables clave que serán utilizadas a lo largo del desarrollo del proyecto. El objetivo es preparar un entorno robusto que permita el análisis eficiente de datos financieros y la implementación de estrategias de *trading* automatizadas.

1. **Variables de Configuración:** Se definen variables como el símbolo del activo (*trading\_symbol*), el tamaño del lote (*lot\_size*), y los parámetros utilizados para los indicadores técnicos (por ejemplo, media móvil simple *SMA\_value*).
2. **Librerías Utilizadas:** Se hace uso de **TA-Lib (Technical Analysis Library)**, una librería que incluye más de 150 funciones para el análisis técnico. Entre las funciones más comunes que se utilizarán en este proyecto están:
  - **SMA (Simple Moving Average):** Media móvil simple para suavizar los precios y detectar tendencias.
  - **RSI (Relative Strength Index):** Indicador de fuerza relativa que mide la velocidad y el cambio de los movimientos de precio.
  - **ADX (Average Directional Index):** Índice que mide la fuerza de una tendencia.
  - **WT (WaveTrend Oscillator):** Oscilador que identifica puntos de reversión de precios.
  - **CCI (Commodity Channel Index):** Indicador que ayuda a identificar ciclos de sobrecompra o sobreventa.
  - **pandas y numpy** para la manipulación y análisis de datos.
  - **matplotlib** para la visualización de datos (gráficos de velas).
3. **Función Principal:** Se define una función llamada *run\_code()* que sirve como punto de entrada del programa, encargada de ejecutar todas las etapas de análisis y *trading* de forma secuencial. Esta función centraliza el flujo de trabajo, permitiendo una ejecución cíclica del código.

### 4.3. Obtención de Datos Históricos

Para llevar a cabo un análisis riguroso del mercado y ejecutar operaciones de *trading*, es esencial disponer de datos de precios históricos precisos y actualizados. En esta sección, se detallan los pasos necesarios para obtener datos históricos del mercado utilizando la API del proveedor de datos “Alpaca Markets”, filtrando y organizando la información en un *DataFrame* para su posterior análisis técnico.

#### 1. Extracción de Datos desde la API:

- Utilizamos una función proporcionada por la API (*api.get\_crypto\_bars*) para obtener **datos de precios históricos** de un activo en particular.
- **trading\_symbol**: El símbolo del activo a analizar, en este caso el índice Bitcoin/dólar estadounidense BTC/USD).
- **timeframe**: Los datos se extraen con una **resolución de 15 minutos** (*timeframe = "15Min"*), lo que es ideal para un análisis detallado del mercado a corto plazo.
- **start y end**: El rango de datos extraído abarca los últimos **60 días**, lo que proporciona un conjunto de datos suficientemente amplio para identificar patrones recientes y tomar decisiones basadas en tendencias actuales.
- Al utilizar un intervalo de 15 minutos durante un período de 60 días, se obtienen **5,760 registros (96 registros por día x 60 días)**.

#### 2. Filtrado y Organización de los Datos:

- Una vez descargados, los datos contienen múltiples columnas, pero para nuestro análisis nos centramos en un conjunto específico de columnas: [*'open', 'high', 'low', 'close', 'volume'*]. Estas columnas proporcionan la información esencial sobre los precios y el volumen de negociación en cada intervalo de tiempo.

### 4.4. Clasificación por Distancia de Lorentz

En esta sección, se realiza la configuración e implementación del modelo de **Clasificación por Distancia de Lorentz**, una técnica diseñada para clasificar y detectar señales en el mercado basadas en indicadores técnicos.

Para determinar los parámetros óptimos del modelo, se emplea un proceso de pruebas de parámetros por separado (ver *Apéndice A. Código completo a ambiente de pruebas - Backtesting Bot V1.0.ipynb*). Durante este proceso, se ajustan de manera iterativa los distintos parámetros hasta alcanzar una configuración que maximice la precisión y la eficiencia del modelo para la toma de decisiones de *trading*.

La función ***LorentzianClassification*** realiza una clasificación utilizando una serie de características técnicas basadas en datos financieros. Su flujo de trabajo se puede dividir en los siguientes pasos:

1. **Cálculo de la Distancia de Lorentz:** En lugar de usar una simple diferencia entre valores (como en la Distancia Euclidiana), la función Lorentziana toma en cuenta una fórmula más compleja para la medición de la distancia entre puntos. La fórmula de la Distancia de Lorentz generalmente se expresa como:

$$d_L(p, q) = \sqrt{-(t_q - t_p)^2 + (x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2}$$

Donde:

- $d_L(p, q)$ : Es la Distancia de Lorentz entre los puntos  $p$  y  $q$ .
- $t_p, t_q$ : Son las coordenadas temporales de los puntos  $p$  y  $q$ , respectivamente.
- $x_p, x_q$ : Son las coordenadas espaciales en el eje  $x$  de los puntos  $p$  y  $q$ .
- $y_p, y_q$ : Son las coordenadas espaciales en el eje  $y$  de los puntos  $p$  y  $q$ .
- $z_p, z_q$ : Son las coordenadas espaciales en el eje  $z$  de los puntos  $p$  y  $q$ .

En el código proporcionado (*ver Apéndice A. Código completo de implementación - V1.0 Alpaca - Paper Bot - SMA Slope.ipynb*), esta fórmula se aplica sobre el conjunto de datos para medir la distancia entre las características de las muestras (SMA, RSI, ADX, WT, CCI) y las etiquetas de clasificación.

2. **Cálculo de Relevancia de los Datos:** La función Lorentziana no solo mide la proximidad, sino que también actúa como un filtro que da mayor peso a los puntos más cercanos, en comparación con los puntos distantes.

En el código, se calcula esta distancia para cada punto de los datos de entrada (el precio de cierre) respecto a los puntos históricos, considerando solo aquellos puntos cuyo valor de distancia esté por debajo de un umbral, lo que ayuda a identificar cuáles son los eventos o datos más influyentes en la decisión final.

3. **Clasificación y Etiquetado de los Puntos:** Después de calcular las Distancias de Lorentz para todos los puntos de datos relevantes, el siguiente paso es clasificar los puntos según su proximidad a los "puntos de referencia" más cercanos, que son los eventos del pasado más significativos en términos de impacto en el mercado o en el comportamiento del activo.
  - **Puntos cercanos:** Aquellos puntos cuya Distancia de Lorentz es baja se consideran más relevantes y, por lo tanto, tienen un mayor peso en la predicción. Dependiendo de la naturaleza del modelo, estos puntos pueden indicar una tendencia alcista o bajista en el mercado.

- **Puntos lejanos:** Los puntos cuya distancia es alta se consideran menos influyentes en el modelo y, por lo general, no se consideran al momento de hacer la clasificación final.
4. **Decisión Final de Clasificación:** Finalmente, el modelo utiliza las distancias calculadas para determinar a qué clase pertenece un punto de datos dado. Este proceso puede implicar la asignación de una etiqueta de clase "compra" o "vende", en función de los patrones de proximidad entre los puntos de datos actuales y los históricos más relevantes.

Ejemplo: Si estamos utilizando un conjunto de datos sobre los precios de las acciones, la función Lorentziana podría clasificar un día específico como una señal para "comprar" si el precio en ese día tiene una distancia pequeña con respecto a otros días en los cuales el precio aumentó significativamente en el futuro.

En resumen, el uso de la función Lorentziana en la clasificación por distancia tiene la ventaja de dar más relevancia a los eventos o puntos de datos que están más cerca en el tiempo o en las características, lo que mejora la capacidad de predicción del modelo, especialmente en contextos donde las relaciones locales son cruciales (como en el análisis de mercados financieros). Este proceso se realiza de forma iterativa sobre los datos, ajustando las decisiones a medida que el modelo evalúa las distancias de todos los puntos involucrados.

## 4.5. Cálculo y Registro de Operaciones

El proceso de Cálculo y Registro de Operaciones tiene como objetivo registrar las operaciones ejecutadas según las señales de *trading* generadas por el modelo y calcular métricas clave que permiten evaluar el rendimiento de cada operación. Las métricas clave incluyen el número de PIPs obtenidos (que refleja la ganancia o pérdida en términos de variación de precio) y el rendimiento de la operación.

Este proceso se encarga de capturar, registrar y calcular las métricas de las operaciones ejecutadas a partir de las señales generadas por el modelo. Para cada operación identificada (ya sea un *trade* largo o corto), el código registra el valor de entrada, calcula las métricas asociadas como la diferencia de regresión del kernel y el número de PIPs, y finalmente almacena toda esta información en un *DataFrame* para su posterior análisis. Esto permite evaluar el desempeño de las operaciones ejecutadas, proporcionando información valiosa sobre el rendimiento del sistema de trading.

### 4.5.1. Descripción Detallada del Proceso

1. **Inicialización de Variables:** Se comienza estableciendo una variable *start\_price* en *None*, que almacenará el precio de inicio de la operación. También se crea un *DataFrame* vacío llamado *result\_df* con columnas que serán utilizadas para registrar la información relevante de cada operación, como el tipo de *trade*, la diferencia de

regresión del kernel, el número de barras mantenidas, si el *trade* fue ganador o no, y el número de PIPs obtenidos.

2. **Iteración a Través de las Filas del Conjunto de Datos:** El ciclo principal recorre cada fila del conjunto de datos, excluyendo la última fila, que generalmente contiene datos incompletos o futuros no ejecutados. Durante esta iteración, se busca identificar las señales de inicio de un *trade*, ya sea de compra (*startLongTrade*) o de venta (*startShortTrade*).
3. **Identificación de la Señal de Entrada:** Para cada fila, se verifica si existe una señal para iniciar un *trade* largo o corto. Si se encuentra una señal (es decir, si el valor de *startLongTrade* o *startShortTrade* no es *NaN*), se asigna un valor correspondiente a *value\_a*, que es el precio de inicio de la operación, y se determina si el *trade* es "Long" (compra) o "Short" (venta).
4. **Cálculo de la Diferencia de Regresión del Kernel (yhat):** Se calcula la diferencia de las predicciones *yhat* entre dos modelos de regresión, *lc.yhat2* y *lc.yhat1*. Esta diferencia representa la variación en las predicciones de precios entre los dos modelos, lo que puede ser útil para comprender la dirección o fuerza de la tendencia del mercado.
5. **Número de Barras Mantenidas:** Se extrae el número de barras (*barsHeld*) de la fila anterior al índice actual (fila *i-1*). Esta información indica cuántas unidades de tiempo (por ejemplo, velas o barras de un gráfico) han transcurrido desde que se abrió la operación hasta el momento actual.
6. **Cálculo de PIPs:** Los PIPs son una medida del cambio en el precio del activo. Si se ha identificado un valor de inicio (*value\_a*), se asigna a *start\_price* el valor encontrado en esa fila, que será utilizado para calcular la variación en PIPs. Si no se ha encontrado un valor de inicio, el precio de inicio se mantiene sin cambios.
7. **Registro de los Resultados de la Operación:** Se crea un diccionario que almacena los valores específicos para la operación actual, tales como:
  - Valor A: El valor de inicio encontrado.
  - Tipo de *trade*: El tipo de operación (*Long* o *Short*).
  - *Kernel Regression Difference*: La diferencia en las predicciones de regresión del kernel.
  - *Bars Held*: El número de barras mantenidas desde la apertura del *trade*.
  - *Winning Trade*: Inicialmente se establece en 0, pero se actualizará más tarde si el *trade* es ganador.
  - PIPs: Inicialmente establecido en 0, se calculará más tarde.
8. **Agregado de los Resultados al DataFrame:** Finalmente, los resultados de la operación, representados por el diccionario, se agregan al *DataFrame result\_df*, que almacena toda la información sobre las operaciones ejecutadas. Esto se realiza

usando *pd.concat* para concatenar las nuevas filas con las ya existentes, manteniendo el índice actualizado.

## 4.6. Cálculo de Métricas de Desempeño

Una vez registradas las operaciones de *trading*, se calculan diversas métricas clave para evaluar el rendimiento general de las operaciones. Las métricas generadas en este proceso incluyen:

### 4.6.1. PIPs Totales por Operación

Los PIPs son la unidad de medida para la variación del precio de un activo financiero. Esta métrica calcula el cambio de precio durante cada operación. Dependiendo del tipo de *trade* (*Long* o *Short*), se calculan los PIPs como la diferencia entre el precio de cierre del *trade* y el precio de inicio.

- En un **trade Long**, los PIPs se calculan como la diferencia entre el precio de cierre (al siguiente cambio de señal) y el precio de apertura (Valor A).
- En un **trade Short**, la fórmula es similar, pero el cálculo se invierte, ya que se espera una caída en el precio.
- Los **PIPs acumulados** son la suma de todos los PIPs obtenidos en las operaciones ejecutadas hasta el momento.

### 4.6.2. Número de Operaciones Ganadoras y Perdedoras

Se registra si una operación fue **ganadora** o **perdedora** comparando el precio de apertura (Valor A) con el precio de cierre (el valor al siguiente cambio de señal). Si el *trade* es **Long** y el precio final es superior al inicial, se marca como **ganadora**; si es **Short** y el precio final es inferior al inicial, también se marca como ganadora. Si no se cumplen estas condiciones, la operación es clasificada como **perdedora**.

- **Operación ganadora:** Se marca con un valor de 1 cuando el *trade* es exitoso.
- **Operación perdedora:** Se marca con un valor de -1 cuando el *trade* no es rentable.

### 4.6.3. Tasa de Éxito (*Win Rate*)

La **tasa de éxito** o **win rate** es el porcentaje de operaciones ganadoras con respecto al total de operaciones ejecutadas. Se calcula como:

$$Win Rate = \left( \frac{\text{Operaciones ganadoras}}{\text{Total de operaciones}} \right) \times 100$$

Esta métrica refleja la efectividad del modelo de *trading* en cuanto al porcentaje de operaciones que resultan en ganancias.

#### 4.6.4. Ratio de Sharpe

El Ratio de Sharpe es una métrica que evalúa el rendimiento ajustado por riesgo de las operaciones. Compara el exceso de rendimiento obtenido con respecto a la tasa libre de riesgo (en este caso, un 2% anual) con la volatilidad de los rendimientos (medida por la desviación estándar).

El Ratio de Sharpe se obtiene dividiendo el exceso de rendimiento por la desviación estándar. Este valor indica cuán eficiente ha sido el modelo en generar rendimientos ajustados por riesgo. El cálculo del Ratio de Sharpe se realiza de la siguiente forma:

$$\text{Ratio de Sharpe} = \frac{\bar{R} - R_f}{\sigma_R}$$

Donde:

- $\bar{R}$  = Promedio de los rendimientos diarios (%).
- $R_f$  = Tasa libre de riesgo anual dividida entre 252 (días hábiles).
- $\sigma_R$  = Desviación estándar de los rendimientos diarios (%).

El Ratio de Sharpe se expresa en porcentaje y se interpreta como un indicador del rendimiento ajustado por riesgo. Un valor más alto indica un mejor rendimiento con menor riesgo.

##### 4.6.4.1. Interpretación de Valores

1. **Ratio de Sharpe Negativo (< 0):** Indica que el rendimiento de la inversión fue menor que la tasa libre de riesgo, lo cual no compensa el riesgo asumido. Este resultado sugiere que el inversor habría estado mejor invirtiendo en un activo libre de riesgo [17].
2. **Cercano a 0:** Implica que el rendimiento adicional generado es prácticamente inexistente o marginal con respecto al riesgo asumido.
3. **Entre 0.5 y 1:** Este rango indica un rendimiento aceptable en relación con el riesgo. Es típico de inversiones moderadamente eficientes.
4. **Mayor a 1:** Se considera un buen Ratio de Sharpe, ya que sugiere que el rendimiento ajustado por riesgo es significativo. Las estrategias con valores en este rango suelen ser atractivas para los inversores.

5. **Mayor a 2:** Representa un rendimiento excelente con un bajo nivel de riesgo en comparación. Este nivel es típico en estrategias o activos altamente eficientes.
6. **Excepcional (> 3):** Aunque poco común, un Ratio de Sharpe tan alto puede ocurrir en períodos o estrategias específicas, y denota un equilibrio extremadamente favorable entre rendimiento y riesgo.

#### 4.6.4.2. Importancia Práctica

El Ratio de Sharpe permite a los inversores:

- Comparar diferentes activos o estrategias en igualdad de condiciones ajustadas por riesgo.
- Identificar si el rendimiento obtenido justifica el nivel de riesgo asumido.
- Elegir entre opciones de inversión, priorizando aquellas que maximizan el rendimiento por unidad de riesgo.

Sin embargo, es importante recordar que el Ratio de Sharpe asume una distribución normal de los rendimientos y puede no capturar correctamente riesgos extremos o eventos inusuales. A pesar de sus limitaciones, sigue siendo una herramienta estándar en análisis financiero [17].

### 4.7. Visualización de Resultados

El código genera un gráfico de velas (*candlestick chart*) diseñado para ofrecer una representación visual detallada de las dinámicas del mercado, integrando múltiples indicadores técnicos y señales de operación. Este tipo de gráfico es ampliamente utilizado en análisis técnico, ya que muestra de manera efectiva los movimientos de precios a lo largo del tiempo, reflejando valores de apertura, cierre, máximos y mínimos para cada período seleccionado.

Además de las velas, el gráfico incorpora elementos clave que ayudan a evaluar las *oportunidades de entrada y salida* en los mercados. Entre estos elementos se encuentran indicadores como la Media Móvil Simple (SMA), señales específicas para estrategias de compra (*startLongTrade*) y venta (*startShortTrade*), y estimaciones basadas en regresión kernel (*lc.yhat1* y *lc.yhat2*). Estos indicadores están superpuestos al gráfico principal, proporcionando un contexto adicional que mejora la interpretación de las señales.

El objetivo principal es ofrecer un medio intuitivo y accesible para analizar el desempeño de las estrategias de *trading* y tomar decisiones informadas basadas en datos visuales claros y estructurados.

## 4.8. Ejecución de Orden de Operación

Finalmente, el código ejecuta **órdenes de compra y venta automáticas** basadas en las señales detectadas por el modelo, asegurando una gestión eficiente y evitando operaciones duplicadas. Este proceso automatizado forma parte de una estrategia de *trading* algorítmico, donde las decisiones se toman de manera programática y en tiempo real.

## 4.9. ¿Qué hace y cómo funciona?

1. **Detección de Nuevas Señales:** El sistema compara el valor actual de la señal (Valor A) con el valor registrado en la iteración anterior (*valor\_a\_anterior*). Si estos valores son diferentes, se interpreta que hay una nueva señal de operación.
2. **Evitar Duplicados:** Antes de ejecutar una nueva orden, el código verifica si el tipo de operación actual (Tipo de *trade*) difiere del anterior. Si ambos tipos son iguales, no se realiza ninguna acción, evitando operaciones redundantes.
3. **Cierre de Posiciones Abiertas:** Cuando se detecta un cambio en el tipo de operación (por ejemplo, de "Long" a "Short"), el sistema liquida todas las posiciones abiertas mediante la función *api.close\_all\_positions*. Esto garantiza que las nuevas órdenes no se superpongan con posiciones previas.
4. **Ejecución de Órdenes:** Una vez validada la señal, el sistema envía una orden de compra o venta al mercado utilizando la función *api.submit\_order*. Los parámetros incluyen:
  - **Símbolo:** El activo que se está operando.
  - **Tipo de operación:** "Buy" para operaciones Long y "Sell" para Short.
  - **Cantidad:** El número de acciones a operar.
  - **Tipo de orden:** Orden a mercado (*market*) para ejecución inmediata.
  - **Duración:** Configurada como "Good Till Canceled" (GTC), lo que significa que la orden permanecerá activa hasta ser ejecutada o cancelada manualmente.
5. **Actualización del Estado:** Finalmente, el valor actual (Valor A) se almacena en la variable global *valor\_a\_anterior*, para que en la próxima iteración se pueda identificar si existe un cambio.

---

## 5. RESULTADOS Y DISCUSIÓN

---

### 5.1. Resultados

El modelo de clasificación basado en Clasificación por Distancia de Lorentz aplicado al par BCH/USD ha demostrado un desempeño notable al analizar los datos históricos y generar señales de *trading*. Durante la fase de prueba, se analizaron datos históricos de precios correspondientes al par dólares por Bitcoin (BCH/USD), abarcando un período de **60 días** y una periodicidad de **15 minutos**, lo que resultó en la recopilación de más de **5,000 registros de datos**. Estos datos permitieron obtener una base sólida para el análisis y entrenamiento del modelo de clasificación mismos que se muestran en la Ilustración 7 y 8 para dos periodos de tiempo distintos respectivamente. A continuación, se presentan los hallazgos más relevantes:

#### 1. Tasa de Éxito:

- Con un total de **70 operaciones ejecutadas**, el modelo logró una tasa de éxito del **92.86%** (65 operaciones ganadoras frente a solo 5 operaciones perdedoras). Esto indica que la estrategia propuesta tiene un alto nivel de precisión en la detección de oportunidades de *trading*, minimizando las pérdidas potenciales.

#### 2. Acumulación de Pips Positiva:

- El sistema acumuló **3056.02 PIPs** a lo largo del período de evaluación. Este resultado refleja un rendimiento sostenido, destacando la eficacia del modelo en la generación de ganancias netas significativas.

#### 3. Optimización y Robustez del Modelo:

- La integración de múltiples **indicadores técnicos optimizados** mediante un proceso de pruebas (Ilustración 9) permitió ajustar los parámetros clave para maximizar el rendimiento del modelo.

#### 4. Evaluación del Riesgo con el Ratio de Sharpe:

- El **Ratio de Sharpe** obtenido fue de **0.13**, lo que indica un equilibrio adecuado entre el retorno y el riesgo asumido. Si bien este valor puede sugerir un riesgo moderado en las operaciones, la alta tasa de éxito compensa potencialmente esta métrica al garantizar un alto porcentaje de operaciones ganadoras.

En general, los resultados demuestran que la estrategia basada en Clasificación por Distancia de Lorentz es altamente efectiva para la identificación de señales de *trading* en el par BCH/USD, proporcionando una ventaja considerable en la toma de decisiones de inversión.

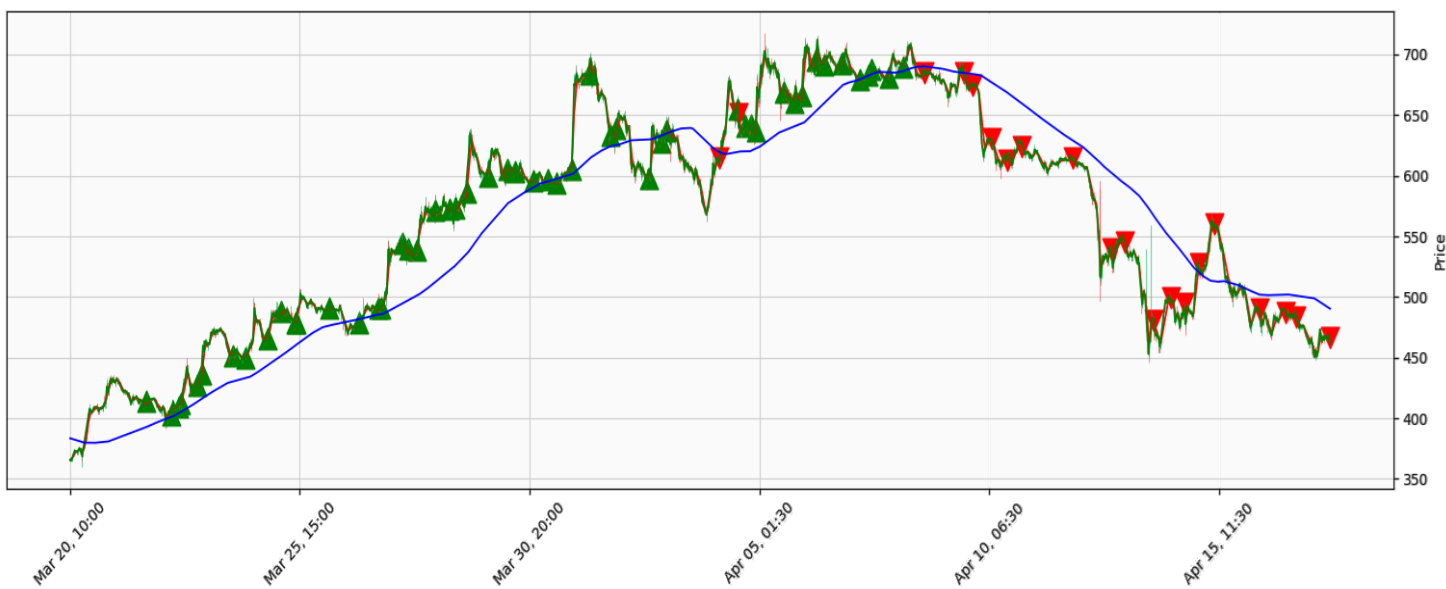


Ilustración 7. Gráfico de Velas del Símbolo BCH/USD durante el periodo de Marzo 20 y Abril 25 del 2024.

- Símbolo: BCH/USD
- Total de operaciones: 70
- Operaciones ganadoras: 65
- Operaciones perdedoras: 5
- **Tasa de éxito: 92.86%**
- Pips acumulados: 3056.02
- Ratio de Sharpe: 0.13

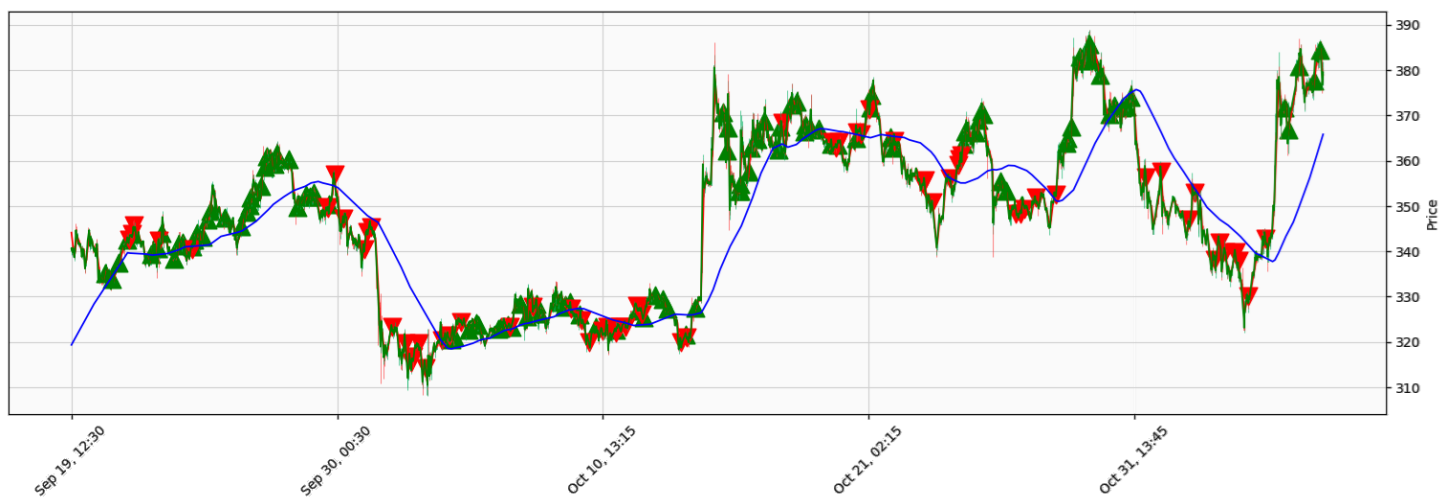


Ilustración 8. Gráfico de Velas del Símbolo BCH/USD durante el periodo de Septiembre 19 y Noviembre 8 del 2024.

- Símbolo: BCH/USD
- Total de operaciones: 182
- Operaciones ganadoras: 154
- Operaciones perdedoras: 28
- **Tasa de éxito: 84.62%**
- Pips acumulados: 673.02
- Ratio de Sharpe: 0.11

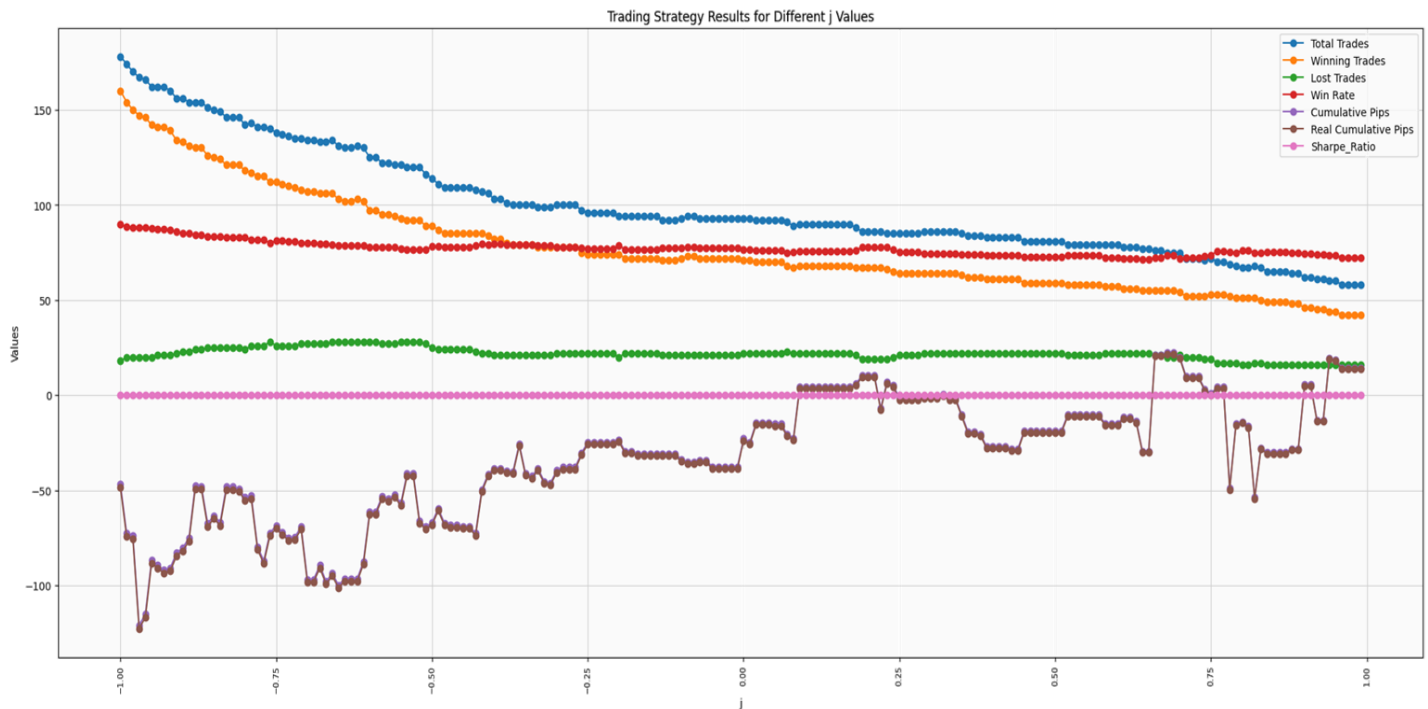


Ilustración 9. Resultados del escenario de pruebas para múltiples valores de ajuste.

## 5.2. Discusión

El análisis y los resultados obtenidos destacan varios aspectos importantes:

- **Desempeño Consistente en Condiciones de Mercado Variables:**
  - La inclusión de filtros de volatilidad y régimen ha demostrado ser efectiva para mejorar la adaptabilidad del modelo, permitiendo que se ajuste dinámicamente a diferentes entornos de mercado. Esto es crucial en el mercado de criptomonedas, donde la **volatilidad es elevada** y las condiciones cambian rápidamente.
- **Impacto del Proceso de Optimización:**
  - La optimización de los indicadores a través del proceso de pruebas permitió ajustar la sensibilidad del modelo, mejorando la precisión de las predicciones.

Sin este proceso iterativo, la tasa de éxito y la acumulación de PIPs habrían sido considerablemente menores.

- **Análisis del Ratio de Sharpe:**
  - Aunque el **Ratio de Sharpe** fue moderado (0.13), la alta **tasa de aciertos del 92.86%** sugiere que, a pesar de un nivel de riesgo presente, la estrategia es capaz de capturar consistentemente movimientos favorables del mercado, minimizando la exposición a operaciones no rentables.
- **Limitaciones y Oportunidades para Futuras Mejoras:**
  - A pesar del alto rendimiento alcanzado, es posible que el modelo se beneficie de ajustes adicionales en los filtros o de la inclusión de **salidas dinámicas** para reducir aún más el riesgo y optimizar el Ratio de Sharpe.
  - Futuras investigaciones podrían explorar el uso de técnicas adicionales como **suavizado de kernel** o incorporar **redes neuronales** para mejorar la capacidad de generalización del modelo en datos no vistos.

---

## 6. CONCLUSIONES

---

El presente estudio demuestra que la aplicación de Ciencia de Datos en el ámbito del *trading* automatizado puede optimizar significativamente la toma de decisiones financieras.

Utilizando algoritmos avanzados como el de Clasificación por Distancia de Lorentz y el de Aproximación por Vecinos Cercanos, se lograron resultados excepcionales en la predicción de movimientos de precios en el par BCH/USD.

**Implementación Exitosa:** Los algoritmos desarrollados se integraron con éxito en una plataforma automatizada, permitiendo el análisis de más de 5,000 datos históricos de precios para obtener señales precisas de compra y venta.

**Señales Prometedoras:** Con una tasa de acierto de **92.86%** (65 operaciones ganadoras de un total de 70), el sistema demostró ser altamente efectivo, acumulando **3,056.02 PIPs** en un período de **60 días**. Esto respalda la fiabilidad del enfoque propuesto y su potencial para reducir las señales falsas y optimizar las decisiones de *trading*.

**Procesamiento Automático Eficiente:** El sistema fue capaz de procesar transacciones de manera automática, minimizando la intervención humana. Esto permitió una operación continua y más eficiente en un entorno de alta volatilidad, mejorando la precisión en las operaciones y maximizando el retorno.

### 6.1. Trabajo Futuro

En última instancia, se propone implementar un modelo de aplicación automática basado en sistemas en la nube (*Webhook*). Este modelo automatizará la ejecución de los algoritmos desarrollados, permitiendo la generación eficiente y objetiva de transacciones eliminando la intervención humana y los recursos locales que deben mantenerse ejecutando los algoritmos, contribuyendo así al avance y la eficacia de las estrategias de *trading* y su automatización. *Ver apéndice A. Código completo de implementación - V1.0 Alpaca - Paper Bot - SMA Slope.ipynb*

---

## 7. BIBLIOGRAFÍA

---

- [1] Z. K. A. & M. A. J. Bodie, *Historia del trading y herramientas financieras tradicionales*, McGraw-Hill Education, 2020.
- [2] E. P. Chan, *Algorithmic Trading: Winning Strategies and Their Rationale*, Wiley, 2013.
- [3] I. B. Y. & C. A. Goodfellow, «Aplicaciones de Aprendizaje Automático en datos financieros,» de *MIT Press*, 2016.
- [4] R. C. C., «¿Que es el FOMC? – Importancia Para Los Mercados Financieros.,» [En línea]. Available: <https://www.tecnicasdetrading.com/2016/09/fomc-mercados-financieros.html>.
- [5] R. G. a. G. E. A. P. A. Batista, «An Empirical Comparison of Dissimilarity Measures for Time Series Classification,» de *2013 Brazilian Conference on Intelligent Systems*, 2013.
- [6] Y. K. a. H. H. U. H. Ş. Bilge, A new classification method by using Lorentzian distance metric, ResearchGate, 2015.
- [7] Y. K. a. H. Ş. Bilge, «Lorentzian Distance Classifier for Multiple Features,» de *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, 2017.
- [8] H. Ş. B. a. H. H. U. Y. Kerimbekov, «The use of Lorentzian distance metric in classification problems,» *Pattern Recognition Letters*, vol. 84, p. 170–176, December 2016.
- [9] V. S. P. e. al., «Effects of Distance Measure Choice on KNN Classifier Performance - A Review,» 2019. [En línea]. Available: <https://arxiv.org/pdf/1708.04321.pdf>.
- [10] A. B. J. L. a. J. L. A. Bagnall, The Great Time Series Classification Bake Off: An Experimental Evaluation of Recently Proposed Algorithms, ResearchGate, 2016.
- [11] J. dehorthy, «Machine Learning: Lorentzian Classification,» February 2023. [En línea]. Available: <https://www.tradingview.com/script/WhBzgfDu-Machine-Learning-Lorentzian-Classification/>.
- [12] J. S. a. K. Johnson, «Application of Lorentzian Distance Metric in Financial Time Series Analysis,» *Journal of Financial Analytics*, vol. 12, nº 3, pp. 45-56, 2018.
- [13] S. M. D. M. N. N. S. S. R. & W. A. Y. Arya, «An optimal algorithm for approximate nearest neighbor searching in fixed dimensions.,» *Journal of the ACM (JACM)*, vol. 45, nº 6, pp. 891-923, 1998.
- [14] H. A. Lorentz, «La teoría de los fenómenos electromagnéticos en cuerpos en movimiento,» *Archivos Neerlandeses de Ciencias Exactas y Naturales*, vol. 25, pp. 363-552, 1900.

- [15] D. M. M. N. S. N. R. S. y. A. Y. W. A. P. Arya, «An optimal algorithm for approximate nearest neighbor searching in fixed dimensions,» *Journal of the ACM*, vol. 45, nº 6, pp. 891-923, 1998.
- [16] R. R. y. M. L. T. Zhang, «BIRCH: an efficient data clustering method for very large databases,» de *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, 1998.
- [17] W. F. Sharpe, «The Sharpe Ratio,» *The Sharpe Ratio. The Journal of Portfolio Management*, vol. 21, nº 1, pp. 49-58, 1994.
- [18] A. M. Téllez, «Geometrías no-Euclidianas - Capítulo IV: Construyendo una Nueva Geometría,» 29 Septiembre 2007. [En línea]. Available: <https://geometrias-no-euclidianas.blogspot.com/2007/09/parte-iv-construyendo-una-nueva.html>. [Último acceso: 26 Noviembre 2024].

## APÉNDICE A. Accesos directos al código implementado.

**Código completo de implementación** - V1.0 Alpaca - Paper Bot - SMA Slope.ipynb:

- <https://colab.research.google.com/drive/12l9QwcgbJCdoioDz5d1QXwOBFOusK-3i?usp=sharing>

**Código completo a ambiente de pruebas** - Backtesting Bot V1.0.ipynb:

- [https://colab.research.google.com/drive/1qW\\_JprMxCH0Vpu1BK5q7Uo0Bn9dXbmgM?usp=sharing](https://colab.research.google.com/drive/1qW_JprMxCH0Vpu1BK5q7Uo0Bn9dXbmgM?usp=sharing)