



Universidad del
Rosario

Escuela de Ingeniería,
Ciencia y Tecnología



MACC
Matemáticas Aplicadas y
Ciencias de la Computación

Machine learning para arbitraje financiero en el mercado de renta variable colombiano

Daniel Eduardo Ramírez

Jaime Augusto Segura

Director: Alexander Caicedo

Co-director: Edgar Jose Andrade

Presentado para obtener el título de
Magíster en Matemáticas aplicadas y ciencias de la computación

Universidad del Rosario

Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Matemáticas aplicadas y ciencias de la computación

Abstract

El desarrollo y la tecnificación de los mercados de capitales en los últimos años ha derivado en una competencia entre los actores del mismo por la búsqueda de oportunidades de inversión mediante el uso de herramientas computacionales veloces, potentes y sofisticadas. El hallazgo de patrones en algunas oportunidades de inversión cuya duración es de fracciones de segundo pero que pueden ocurrir un sin número de veces en el término de un día, multiplica las oportunidades de aquellos inversionistas que se encuentran bien equipados para explotarlas a su favor. En el presente trabajo, mostraremos cómo se pueden aplicar algunas técnicas de inteligencia artificial para construir estrategias rentables de trading algorítmico en el mercado de renta variable colombiano. Construiremos varios modelos de Machine y Deep Learnig capaces de predecir con precisión aceptable, algunas oportunidades de inversión que se presentan en ventanas cortas de tiempo. Mostraremos con detalle cuáles son las capacidades predictivas de los modelos desarrollados y los retornos esperados.

Palabras clave: Arbitraje, Trading de Alta Frecuencia, Inteligencia Artificial.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	3
2	OBJETIVOS	5
2.1	Objetivo general	5
2.2	Objetivos específicos	5
3	PROBLEMA Y JUSTIFICACIÓN	6
4	MARCO TEÓRICO Y ESTADO DEL ARTE	8
5	METODOLOGÍA	13
5.1	Spread Bid/Ask	14
5.2	Función de Arbitraje	16
5.3	Base de datos	17
5.4	Consideraciones	18
5.5	Aproximación al problema	19
6	RESULTADOS Y DISCUSIÓN	20
6.1	Exploración I	20
6.1.1	Conclusiones exploración	24
6.2	Exploración II	25
6.2.1	Modelo Base de Referencia	26
6.2.2	Otros modelos	27
6.2.3	Selección de Features	28
6.2.4	Variables Categóricas	30
6.2.5	Conclusiones Exploración	30

6.3	Metodología Formal	31
6.3.1	Evaluación para un rango de lags	32
6.3.2	Modelo de mayoría de votos con rango de lags óptimo (MVC)	33
6.3.3	Modelos de evaluación a Futuro	35
6.4	Modelos de Deep Learning	36
6.4.1	Arquitecturas	37
6.4.2	Resultados arquitectura Embeddings	37
6.4.3	Resultados Arquitectura capas densas	38
6.4.4	Resultados arquitectura convolucional	39
6.4.5	Métricas de desempeño Deep Learning	40
6.5	Evaluación Financiera	41
6.5.1	Filtro periodo exdividendo	42
6.5.2	Tratamiento días festivos	42
6.5.3	Resultados sesión abierta con subastas	43
6.5.4	Pérdidas y Ganancias	44
6.5.5	Pérdidas y ganancias incluyendo días festivos	45

7 CONCLUSIONES 47

1 INTRODUCCIÓN

En finanzas el termino arbitraje se utiliza para denominar un flujo de caja sintético, por lo general una operación simultánea de compra/venta de uno o varios instrumentos financieros cuyo resultado en términos de utilidad es positivo. Esta utilidad se logra sin asumir riesgos en la ejecución de las operaciones por lo que la

misma está garantizada a priori.

El presente proyecto tiene como propósito desarrollar un modelo que determine a priori y de forma sistemática las oportunidades de arbitraje de pares para las acciones colombianas doblemente listadas, de modo que se puedan aprovechar las diferencias de precio de un activo que se cotiza simultáneamente en dos o más mercados. Estas oportunidades ocurren gracias a las combinaciones de precios de las acciones en el mercado local y las bolsas extranjeras en las que el instrumento está listado, junto al precio de la tasa de cambio USD/COP en un momento determinado. Es importante anotar que este tipo de oportunidades de inversión ocurre en ventanas de tiempo muy cortas: fracciones de segundos a minutos.

Un ejemplo de lo anterior puede ocurrir con una acción como la de Ecopetrol, cuyo mercado de origen es Colombia, pero que tiene un doble listado con ADRs (American Depositary Receipt) que cotizan en la bolsa de Nueva York. Ocurre frecuentemente que al equiparar el precio de cotización del instrumento en una misma moneda, por ejemplo, el dólar americano, hay diferencias instantáneas en los precios de cotización entre uno y otro mercado [1]. En este caso, hay una operación de arbitraje cuando se compra el activo en el mercado cuyo precio es menor y simultáneamente se realiza la venta en el mercado dual: la utilidad se obtiene de la liquidación en una misma moneda de ambas operaciones.

El objetivo general del proyecto consiste en encontrar un modelo que logre predecir, con suficiente precisión, el momento en que se van a presentar condiciones como la anteriormente descrita para con base en esto construir una estrategia de trading rentable que se ajuste a las características operativas de los mercados involucrados.

2 OBJETIVOS

2.1 Objetivo general

- Desarrollar un modelo rentable y que se ajuste a las características operativas de los mercados donde se implemente que permita explotar ineficiencias en los precios de las acciones doblemente listadas mediante la predicción de oportunidades de arbitraje de corto plazo.

2.2 Objetivos específicos

- Revisar distintos modelos y técnicas de Machine y Deep Learning para evaluar cual ofrece el mejor desempeño en el contexto del problema.
- Revisar distintas aproximaciones para la solución del problema (Clasificación multiclase, clasificación binaria, clasificación en un único tipo de arbitraje)
- Determinar los rezagos óptimos (cuantos precios hacia atrás hay que mostrarle al modelo) que permitan alcanzar las mejores métricas en la predicción de los eventos de arbitraje.
- Determinar el horizonte de predicción que permita alcanzar las mejores métricas en la predicción de los eventos de arbitraje a futuro.
- Realizar una evaluación monetaria del mejor modelo para determinar su utilidad al implementarlo como una estrategia de trading en un entorno real.

3 PROBLEMA Y JUSTIFICACIÓN

Durante la segunda mitad del siglo XX la teoría de portafolios se sostuvo sobre la idea de que los mercados financieros son eficientes [2], es decir, que los precios de los activos financieros reflejan toda la información disponible para tomar decisiones de inversión (Hipótesis de mercados eficientes - HME)[3]. El cumplimiento de la HME implica que ningún agente puede acceder con mayor información al mercado, explotar esa posición privilegiada y tomar ventaja de los demás. En términos generales, según la HME, las condiciones en las cuales los operadores acceden a los mercados son iguales para todos.

Una de las implicaciones de tener precios que incorporan toda la información sobre un instrumento financiero y de tener agentes que tienen acceso a esta información en igualdad de condiciones, tal como sugiere la HME, es que no existe arbitraje y es imposible generar retornos por encima del promedio del mercado.

Sin embargo, en la práctica la HME no se cumple como ha sido ampliamente documentado por [4],[5], [6] entre otros. Los mercados financieros se conforman mediante interacciones de inversionistas muy diversos que planean sus estrategias de inversión con horizontes de tiempo distintos: desde administradores de fondos de pensiones que toman posiciones de largo plazo (años), hasta inversionistas con estrategias muy sofisticadas de corto plazo (Analistas cuantitativos - quants) que buscan tomar provecho de los movimientos de los precios en fracciones de segundos. La información con la que estos agentes alimentan sus estrategias de inversión es muy diversa. Existen por ejemplo analistas fundamentales que toman sus decisiones basadas en información macroeconómica, balances de las compañías, estados financieros, etc. y cuya frecuencia de op-

eración es relativamente baja [7]. Por otro lado, están los traders más dinámicos que prefieren el análisis técnico: este grupo opta por monitorear indicadores de tendencias, gráficos e información estadística propia del comportamiento del instrumento para así tomar posiciones con unas frecuencias altas en ventanas cortas de tiempo [8]. También están los inversores ocasionales (mom and pop investors) que llegan al mercado por alguna transacción particular pero cuyo conocimiento sobre la dinámica de este es muy limitado y operan con muy poca o nula información.

La interacción de todos estos agentes tan diversos lleva a que frecuentemente ocurran ineficiencias en el mercado (activos cuyo precio no refleja su verdadero valor). Parte de la dinámica del mercado de capitales está dada por agentes en busca de este tipo de oportunidades las cuales se dan por lo general en ventanas de tiempo muy cortas y requieren de un monitoreo activo y constante del mercado. Encontrar, anticipar y aprovechar estas oportunidades permite generar utilidades libres de riesgo y por encima del promedio del mercado.

Los altos niveles de tecnificación en los mercados de capitales hacen que las oportunidades libres de riesgo sean cada vez más difíciles de detectar y explotar. Tanto el enfoque puramente financiero como el puramente estadístico ya no son suficientes para abordar el problema. Es por esta razón que hoy en día se buscan herramientas más sofisticadas para la detección de los patrones que derivan en estas oportunidades. Las finanzas cuantitativas han encontrado en el Machine Learning, el Deep Learning y la inteligencia artificial unos importantes instrumentos para facilitar dicha tarea [9].

La mayoría de estas estrategias han sido implementadas durante los últimos treinta años en mercados de capitales altamente desarrollados y cuyas características

en términos de volumen, liquidez y cantidad de activos distan bastante del mercado colombiano. No obstante, dado el crecimiento en la aplicación de tecnología financiera en los últimos años, el mercado de renta variable colombiano y en general los mercados de renta variable en el mundo han experimentado una tendencia creciente en el número de operaciones algorítmicas y tecnificadas. El proyecto pretende la creación de una estrategia de arbitraje basada en técnicas de Machine y Deep Learning enfocada en el mercado de renta variable colombiano, cuyo desafío implica encontrar un modelo rentable en un mercado pequeño y en desarrollo, con volúmenes de negociación y liquidez bajos. Sin embargo, por experiencia se sabe que la frecuencia de ocurrencia de este tipo de desviaciones e ineficiencias en mercados pequeños es más alta, por lo que se espera que lograr predecir su ocurrencia con una precisión aceptable.

4 MARCO TEÓRICO Y ESTADO DEL ARTE

El área en la cual se desarrolla este proyecto hace parte de las finanzas cuantitativas. El análisis cuantitativo para finanzas es un área muy extensa que tuvo su mayor desarrollo a finales del siglo XX y principios del siglo XXI y surgió con el propósito de aplicar matemáticas avanzadas y física a la modelación de problemas complejos como la dinámica de los precios de los commodities, las acciones y los bonos, o la valoración de instrumentos derivados. [10]

El análisis cuantitativo ha ocasionado grandes avances en el sector de las finanzas, por ejemplo, el desarrollo electrónico de los mercados financieros que tuvo su mayor auge a finales de los años 90's y principios del siglo XXI apalancado por el apoyo de académicos y facultades de matemáticas y ciencias de las princi-

pales universidades de Estados Unidos. Producto de ello surgió un nuevo perfil profesional en el mercado de capitales, el quant, quien cuenta con una extensa formación especializada en matemáticas, programación y finanzas. Sobre quants y el desarrollo electrónico de los mercados de capitales la investigación parte de dos trabajos que documentan en detalle el desarrollo de las finanzas cuantitativas en los mercados de Estados Unidos: [11] en los que se narra minuciosamente el desarrollo de los mercados electrónicos en el país, desde sus inicios hasta el surgimiento del trading algorítmico. Además, [12] muestra cómo se han desarrollado los mercados de capitales durante los últimos 30 años alrededor de esta figura de quants y el impacto que esta actividad ha traído para la economía global: a los quants se les atribuye la estructuración de los complejos productos financieros que ocasionaron la crisis de los mercados hipotecarios en 2008.

Los quants también impulsaron el rápido desarrollo del trading algorítmico durante la primera década del siglo XXI en EE.UU. El objetivo de este tipo de trading es utilizar códigos y software informáticos para abrir y cerrar posiciones de compra y venta de títulos valores de acuerdo con reglas establecidas, como por ejemplo los movimientos de precio en un mercado subyacente. A medida que los mercados se han desarrollado la competencia por la velocidad en la ejecución ha conducido al trading algorítmico a evolucionar hacia el High Frequency Trading (HFT), u operación de alta frecuencia, cuyo objetivo consiste en abrir y cerrar posiciones de compra y/o venta en fracciones de segundo. La tipología del HFT está dividida por lo menos en los siguientes ítems:

1. Estrategias basadas en proporcionar liquidez.
2. Operativa basada en el análisis de los ticks. (Filter trading y Momentum Trading)

3. Trading estadístico (Arbitraje estadístico y trading técnico)

4. Arbitraje de Latencias

El desarrollo de este proyecto se enmarca dentro de la tercera tipología: Arbitraje estadístico y trading técnico, la cual, por condiciones de volumen y liquidez es más susceptible de ser aplicada en el entorno del mercado colombiano. Adicionalmente, el mercado colombiano réplica más o menos el camino de desarrollo de los mercados evolucionados en términos del trading algorítmico, desde hace algunos años recibe mediante su protocolo de comunicación Fix enrutamiento de órdenes hacia el mercado de renta variable lo que permite implementar este tipo de estrategias en un ambiente real.

Sobre arbitraje estadístico existe una variedad amplia de enfoques: desde cointegración [13], estocasticidad [14], enfoque de series de tiempo [15], hasta sofisticadas herramientas de Machine Learning. Sobre este tema, se toma como punto de partida el artículo de [16] en el que dividen el proceso de arbitraje en dos etapas: la selección de activos y la toma de una decisión sobre cuándo y cómo operar estos activos. En términos generales, se propone una estrategia de arbitraje que busca identificar momentos en los que el spread de una serie de tiempo debería cambiar de dirección para converger hacia su media de largo plazo de tal manera que operar en la dirección contraria a la que se dirige el spread resulta potencialmente beneficioso. Estos momentos los identifican a partir de propiedades estadísticas de la serie y con ayuda de cuadros de análisis técnico (Cuadros Renko y Kagi)

En cuanto al uso de técnicas de machine learning e inteligencia artificial en el mundo financiero; en [17] se estudian y proponen modelos para realizar trading basados en aprendizaje reforzado, en los cuales, por medio de una red recurrente

se ofrece evidencia del aprendizaje de estrategias de trading por parte de un algoritmo que además logra resultados positivos para trading intradiario en el mercado de divisas. Un aspecto a resaltar del artículo es que toma en cuenta los costos transaccionales para ajustar rentabilidades, lo cual resulta muy adecuado de cara a que los resultados tengan un uso práctico. Otros ejemplos de aplicaciones en el sector incluyen a [18], quienes desarrollan un sistema para tomar decisiones de trading basado en árboles de decisión cuyo fin es predecir operaciones que tienen la mayor probabilidad de generar utilidad. Por otro lado, en [19] aplican aprendizaje reforzado para entender como la volatilidad de los retornos de una cuenta de retiro afecta la tasa de ahorro de los individuos en el tiempo. Finalmente, en [20] proponen una solución para encontrar factores relevantes para modelos APT (arbitrage pricing theory) por medio de machine learning.

Centrado en la predicción de precios de acciones [21] reconoce las dificultades que esta tarea implica y hace una comparación de los resultados que se obtienen al realizar predicciones con el análisis tradicional de series de tiempo versus algunos enfoques de aprendizaje automático como redes neuronales y máquinas de soporte vectorial, estos últimos muestran unos resultados significativamente superiores a la hora de realizar predicciones sobre mercados de renta variable tanto para tareas de regresión como de clasificación. Por otra parte, el auge de las criptomonedas ha facilitado también el desarrollo de este tipo de ejercicios de predicción sobre estos activos; en [22] se analiza qué tan predecible es el precio del bitcoin en horizontes de 1 a 60 minutos usando diferentes modelos de machine learning. Estos modelos también han sido objeto de estudio en activos más complejos como los instrumentos derivados; en [23] se evalúan algoritmos de inteligencia artificial y machine learning para comparar su desempeño contra el de un humano en el

mercado de futuros.

La enorme cantidad de información disponible en el sector financiero plantea una oportunidad pero a su vez un desafío para el entrenamiento de modelos de machine learning [24], resulta fundamental escoger la cantidad óptima de información que se le muestra a un modelo para reducir el ruido que se genera y resulta también importante considerar como los diferentes algoritmos manejan ese ruido. El presente proyecto tiene como uno de sus objetivos específicos identificar esa cantidad de información óptima de cara a lograr mejores y más confiables predicciones.

En cuanto al presente proyecto, se toma como punto de partida el artículo [25]. En el desarrollo del trabajo se hace una revisión exhaustiva de todo el proceso para la construcción de un sistema de arbitraje tomando como base el comportamiento de los resultados obtenidos para algunas estrategias de high frequency trading en mercados spot y de derivados chinos: desde la selección de activos hasta la evaluación de la rentabilidad. Inicialmente, se propone un problema de clasificación binaria y multiclase para las condiciones de arbitraje partiendo de unos modelos base (regresión, árboles de decisión y maquinas de soporte vectorial) para la evaluación del mismo. Luego, en una segunda etapa, se construyen modelos ensemble para optimizar las métricas de desempeño. Finalmente se acude a modelos de aprendizaje profundo con el propósito de contrastar los resultados y de observar con más detalle la incidencia de algunas variables categóricas. En cada etapa, se calculan distintas métricas para evaluar el desempeño de los modelos para finalmente concluir con una evaluación financiera de los resultados obtenidos.

5 METODOLOGÍA

El primer paso para desarrollar este proyecto consistió en la creación de una base de datos que contiene la información minuto a minuto desde el 3 de enero de 2022 hasta el 12 de mayo de 2022 de los precios de los activos de interés; puntualmente se tienen los mejores Bid y Ask (puntas de compra-venta) y el precio de cierre de la acción preferencial de Bancolombia que cotiza en la bolsa de valores de Colombia, el ADR de la misma acción que cotiza en EEUU, la tasa de cambio USD/COP, así como los mejores precios disponibles en el libro de ofertas para esta variable en el sistema de negociación Set Fx (el principal sistema de negociación de referencia para el mercado de capitales colombiano). Así mismo se incluyó el valor del índice de industriales Dow Jones como una variable exógena cuyo comportamiento marca una referencia global para el resto de los mercados. Adicionalmente, se construyeron variables booleanas para capturar particularidades de las sesiones de negociación de cada mercado como lo son: días festivos, horarios de negociación verano/invierno, horarios de apertura y cierre, operaciones next day (operaciones de compra-venta de USD/COP realizadas luego de la 1 pm y hasta el cierre del mercado) y operaciones de subasta (operaciones de compra-venta de acciones en el mercado local realizadas en los últimos 5 minutos de negociación de cada día).

En este punto se pueden distinguir dos tipologías de arbitraje:

1. Compra en el extranjero y venta local (BE/SL): comprar el ADR en el mercado de Nueva York y vender la cantidad equivalente de acciones en el mercado local.
2. Compra local y venta en extranjero (BL/SE): comprar acciones en el mer-

cado local y vender la cantidad equivalente de unidades ADR en el mercado de Nueva York.

El arbitraje de precio, tal como se mencionó anteriormente, ocurre por ineficiencias e interacciones que se producen entre agentes diversos. Particularidades específicas de cada mercado tales como operaciones de carry trade por el diferencial de tasas de interés que se observan entre ambos mercados, flujos puntuales de compra o venta de grandes bloques de títulos por parte de inversionistas que participan en el mercado propio de cada país (por ejemplo hedge funds o administradoras de fondos de pensiones locales), movimientos ocasionados por grandes flujos de operaciones cambiarias, rebalances de fondos indexados, etc., pueden causar estas diferencias instantáneas u ocasionales entre las cotizaciones. Dependiendo de alguno de los factores mencionados, ocurre alguna de las tipologías de arbitraje en cuestión y los arbitrajeros tienden a corregirlas en fracciones de tiempo cada vez más cortas.

5.1 Spread Bid/Ask

A partir de la distinción de las dos formas de arbitraje, se agrega a la base de datos dos columnas con los resultados minuto a minuto del spread; diferencia de precios entre la mejor intención de compra (bid) y la mejor intención de venta en el otro mercado (ask) comparados en la misma moneda base (COP). El spread de precios entre uno y otro mercado es la variable que determina, en el mayor de los casos, la ocurrencia de las condiciones de arbitraje. Algo importante a anotar en este punto es que el spread contempla los costos transaccionales más altos en los que incurriría una entidad financiera colombiana para ejecutar estas operaciones durante el periodo observado:

- Costo ADR: 0.008 USD por operación de compra o venta de ADRs en el exterior
- Costo SWAP: 0.5 COP por operación de compra o venta de USDs en el mercado local

Para cada observación de la base, el spread se calculó para cada tipo de arbitraje asumiendo la compra y venta de 1 unidad en cada mercado de la siguiente forma:

$$Spread_{BE/SL} = Bid_{pfb} - (Ask_{adr} + Costo_{ADR}) * (Ask_{usd} + Costo_{SWAP})$$

$$Spread_{BL/SE} = (Bid_{adr} - Costo_{ADR}) * (Bid_{usd} - Costo_{SWAP}) - Ask_{pfb}$$

Donde $Spread_{BE/SL}$ corresponde al spread de precio para compra externa y venta local, $Spread_{BL/SE}$ representa el spread de precio para compra local y venta externa, Bid corresponde al mejor precio al que se puede vender un activo, Ask corresponde al mejor precio al que se puede comprar un activo (mejores intenciones en el libro de ofertas). Los subíndices indican el activo en cuestión de acuerdo con:

- pfb : Acción preferencial de Bancolombia negociada en el mercado local
- adr : ADR sobre la acción de Bancolombia negociada en el mercado extranjero
- usd : Tasa de cambio USD/COP

En la Figura 1 se puede observar la distribución de los spreads de precio para ambas tipologías de arbitraje. En esencia, el objetivo es desarrollar un modelo

capaz de predecir cuando este spread se hace positivo (parte positiva de la distribución).

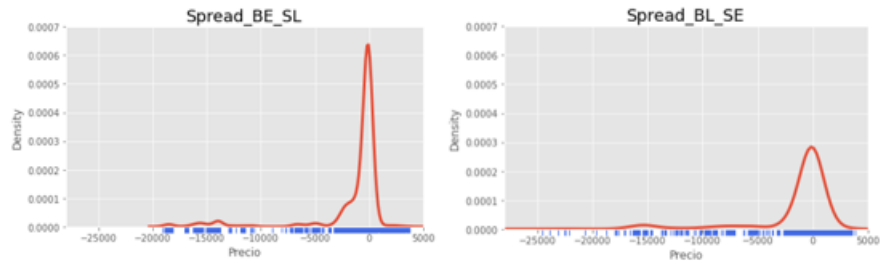


Figura 1: Distribución de los spreads en el periodo observado.

5.2 Función de Arbitraje

Teniendo en cuenta el spread de precio para cada tipo de arbitraje y las condiciones operativas para cada uno de los mercados involucrados, se agregaron a la base de datos dos columnas que indican observación por observación si es posible realizar o no una operación de arbitraje. Para que en un momento t sea posible realizar una operación de arbitraje, es condición necesaria pero no suficiente que el spread de precio entre mercados en el momento t sea positivo, o superior a un determinado umbral (en nuestro caso corresponde a COP 20, umbral que cubre los costos mínimos bursátiles y financieros asociados a la operación durante el periodo observado). Lo anterior puede ocurrir por varias situaciones, como días festivos o subastas (evento propio del mercado de acciones local); eventos en los que a pesar de que teóricamente puede existir un spread positivo para algún tipo de arbitraje, el mismo no se puede materializar simultáneamente por condiciones operativas. Para incorporar estas condiciones se creó una función que evalúa todas

las particularidades que pueden conducir a que no sea posible realizar una operación de arbitraje de la forma BE/SL o BL/SE: esta función tiene en cuenta que los mercados estén abiertos, los festivos locales y del extranjero, las subastas del mercado local y las diferencias de horario que se presentan en algunos momentos del año. Para cada condición, se determinó mediante una función exactamente cuándo y cómo se presenta arbitraje y por qué.

5.3 Base de datos

A partir de este punto ya se cuenta la base definitiva que será el insumo de todos los modelos a desarrollar en adelante, en la Figura 2 se puede observar un extracto de la base.

Date	Hora_COL	Minuto	Hora_NY	Spread_BE_ SL	Spread_BL_ SE	Arbitraje_BE _SL	Arbitraje_BL _SE
2022-01-03	9	54	9	-352,3238	-186,30125	0	0
2022-01-03	9	55	9	-75,37175	-157,96754	0	0
2022-01-03	9	56	9	-85,563	-137,60644	0	0
2022-01-03	9	57	9	-77,621	-137,60644	0	0
2022-01-03	9	58	9	-77,621	-122,9355	0	0
2022-01-03	9	59	9	-69,2819	-122,9355	0	0
2022-01-03	10	0	10	-61,737	-94,09	0	0
2022-01-03	10	1	10	-237,10425	-12,865	0	0
2022-01-03	10	2	10	-258,0317	-25,2895	0	0
2022-01-03	10	3	10	-264,9625	-29,25225	0	0
2022-01-03	10	4	10	-216,275	33,96725	0	1

Figura 2: Extracto base de datos

Las columnas *Spread_BE_SL* y *Spread_BL_SE* corresponden a los spreads de precio para cada tipo de arbitraje (en pesos colombianos) mencionados anteriormente, mientras que las columnas booleanas *Arbitraje_BE_SL* y *Arbitraje_BL_SE* indican si dadas las condiciones y particularidades del mercado es posible realizar

arbitraje la forma BE/SL o BL/SE, respectivamente, en cada minuto.

5.4 Consideraciones

Predecir las oportunidades de arbitraje definidas en la función presenta varias dificultades:

1. Las condiciones actuales del mercado solo nos permiten establecer con certeza si hay una oportunidad de arbitraje en el momento actual: basta con calcular los spreads de precio actuales y conocer las condiciones actuales de cada mercado. Sin embargo, para predecir una oportunidad de arbitraje, hay que establecer qué parámetros del mercado hay que observar hacia el futuro y cuál es el horizonte de tiempo que debemos considerar.
2. El objetivo de predecir en el minuto $t = 0$ una oportunidad de arbitraje para el minuto $t = n$ donde $n > 0$ plantea un desafío en términos de encontrar el valor óptimo de n . Esto significa encontrar un horizonte de predicción ideal (en minutos), ventana de tiempo en la cual un modelo debe establecer si dadas las condiciones de mercado en $t = 0$, se puede inferir que ocurrirá un escenario de arbitraje dentro de n minutos en el futuro. En la Figura 2 esto es equivalente a desplazar hacia arriba n posiciones los valores de las columnas *Arbitraje_BE_SL* y *Arbitraje_BL_SE* y usar esos nuevos valores como objetivo de los modelos a entrenar.
3. El entrenamiento de los modelos se realiza principalmente a partir de información de precios. Para cada activo de interés, el modelo deberá conocer información rezagada acerca de los precios del mismo activo en el pasado. El desafío consiste en encontrar el número de información rezagada óptima

que debe conocer el modelo para lograr buenas predicciones sin caer en problemas de sobre ajuste.

4. El problema se puede abordar de varias maneras: un problema de clasificación para predecir la ocurrencia de solo un tipo de arbitraje (bien sea del tipo BE/SL o bien del tipo BL/SE), un problema de clasificación para predecir la ocurrencia de cualquier tipo de arbitraje, o un problema de clasificación multiclase, para predecir no solo la ocurrencia del arbitraje, sino también el tipo de arbitraje.

5.5 Aproximación al problema

Para solucionar los desafíos anteriormente descritos se proponen varias metodologías de exploración:

1. Exploración I: Modelos de clasificación básicos (Regresión logística, Maquinas de soporte vectorial y árboles de decisión) con combinaciones de rezagos de información desde 1 a 10 y horizontes de predicción de 1 a 10 minutos.
2. Exploración II: Modelos de ensemble (Random Forest), métodos de Boosting (Adaptative Boosting, Gradient Boosting), modelos de última generación (Histogram Gradient Boosting, LightGBM, CatBoost) con 3, 5 y 10 rezagos de información con horizonte de 1 minuto.
3. Metodología Formal: A partir de los resultados de exploración, se construye una metodología formal para evaluar los modelos con mejores resultados, se define una forma para obtener el número de rezagos y el horizonte de tiempo óptimo de predicción así como una función de evaluación financiera

para el mejor modelo y se obtienen los resultados en términos de pérdidas y ganancias.

4. Modelos de Deep Learning: Se aplican técnicas de aprendizaje profundo para incorporar las variables categóricas al análisis y se comparan los resultados con los obtenidos bajo modelos de clasificación de Machine Learning.

6 RESULTADOS Y DISCUSIÓN

6.1 Exploración I

El objetivo principal de la primera exploración consiste en utilizar modelos de baja carga computacional para encontrar combinaciones de horizontes de predicción, así como la cantidad de información rezagada que consume el modelo para su entrenamiento, que conduce a los mejores resultados de pronósticos de oportunidades de arbitraje. Esas combinaciones permiten, en primer lugar, constituirse como base para el entrenamiento de modelos más avanzados y que requieren una carga computacional mayor, y en segundo lugar, obtener unas métricas de rendimiento de referencia.

Los horizontes de predicción a evaluar van desde 1 hasta 10 minutos. Para cada uno de los casos anteriores se entrenan modelos de regresión logística, árboles de decisión y máquinas de soporte vectorial con variables independientes que incluyen las variables categóricas de la base (variables para describir particularidades del mercado) e información de precios de cierre rezagada (acción Preferencial Bancolombia, ADR Bancolombia, Tasa de cambio USD/COP, Spreads) hasta 10 observaciones. Para la evaluación de estos modelos se observa principalmente

el *F1-score* promedio de las clases y la desviación estándar de este puntaje sobre la predicción para una porción de los datos no utilizados durante el entrenamiento, ambos indicadores se obtienen a partir de validación cruzada de cada modelo.

Se utiliza el *F1-score* ya que en el contexto del problema el *accuracy* refleja el desbalance de las clases (la mayoría del tiempo no se presentan oportunidades de arbitraje). La *precisión* y el *recall* son importantes pero resulta difícil definir cual lo es más; un buen algoritmo debe predecir de manera correcta la mayor cantidad de observaciones, pues perderá dinero en la medida que no lo logre, pero también debe capturar la mayor cantidad de oportunidades pues estas resultan escasas. Con ese contexto, consideramos el *F1-score* como la mejor alternativa para la evaluación de los modelos en esta primera etapa de exploración.

El problema anterior puede abordarse de cuatro maneras distintas, a continuación algunos comentarios y resultados de cada una:

- **Predicción de arbitraje sin importar su tipo:** Todos los modelos encuentran que el mejor horizonte de predicción es a 1 minuto pero no resulta claro con ninguno cuanto información rezagada es óptima. Este acercamiento termina siendo de poca utilidad en términos prácticos pues con sus predicciones no es posible tomar una postura en el mercado. ¹
- **Predicción de arbitraje del tipo 1 exclusivamente y predicción de arbitraje del tipo 2 exclusivamente:** Todos los modelos encuentran que el mejor horizonte de predicción es a 1 minuto pero no resulta claro cuanto información rezagada es óptima. Los resultados del tipo 2 resultan ser muy inferiores en regresión logística y ligeramente por debajo en árboles de de-

¹Los resultados detallados pueden consultarse en el Anexo 1 del presente documento el cual esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

cisión y máquinas de soporte vectorial. Este acercamiento si bien puede implementarse en un entorno práctico, requiere de dos modelos de ejecución en simultánea para predecir ambos tipos de arbitraje. Esta dificultad adicional no se ve compensada por un rendimiento particularmente superior al de un modelo de clasificación multiclase.²

- **Predicción de arbitraje del tipo 1 y 2 (Multiclase):** Este acercamiento resulta el más adecuado para llevar a la práctica y dado que los resultados de esta primera exploración resultan positivos, se revisaran en detalle y se fija como la mejor opción para abordar el problema.

En la Figura 3 se pueden observar los resultados de la regresión logística donde el *F1-score* disminuye en la medida que aumenta el horizonte de arbitraje. Por otro lado, el mismo indicador presenta en general ligeros aumentos ($< 1\%$) en la medida que el modelo se entrena con mayor información de precios rezagada. Los *F1-score* en su gran mayoría tienen bajas desviaciones estándar, no obstante los puntajes son discretos. A pesar de lo anterior, la regresión logística obtiene en general mejores resultados con este acercamiento al problema que con los demás. La *precisión* se ubica en rangos entre 76% y 84% pero sufre bastante con el *recall* que apenas llega a un máximo de 50%.

En la Figura 4 se pueden observar los resultados de la clasificación al utilizar árboles de decisión; se observa una mejora notable en los *F1-score* comparados con la regresión logística que adicionalmente muestran menores niveles de desviación estándar. Por otro lado, se mantiene el mismo patrón

²Los resultados detallados pueden consultarse en el Anexo 1 del presente documento el cual esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

donde el *F1-score* disminuye en la medida que aumenta el horizonte de arbitraje, en cuanto a la cantidad de información rezagada el resultado es contrario a lo que mostraba la regresión logística; en la medida que el modelo se entrena con mayor información de precios rezagada el *F1-score* tiende a disminuir levemente.

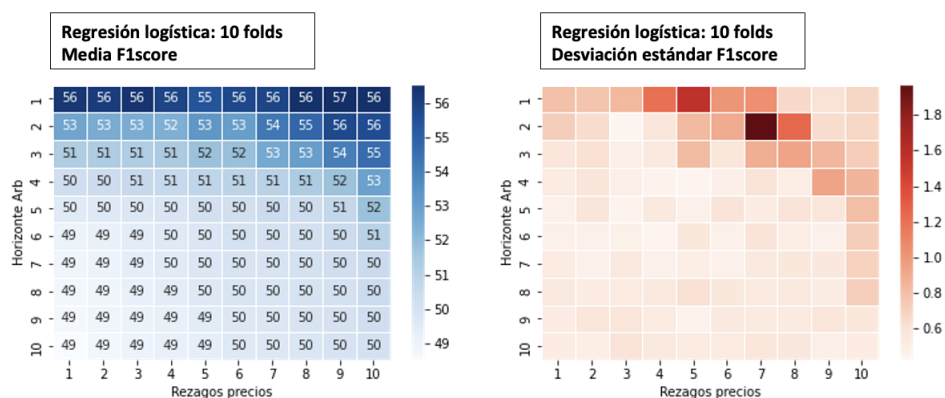


Figura 3: Resultados regresión logística

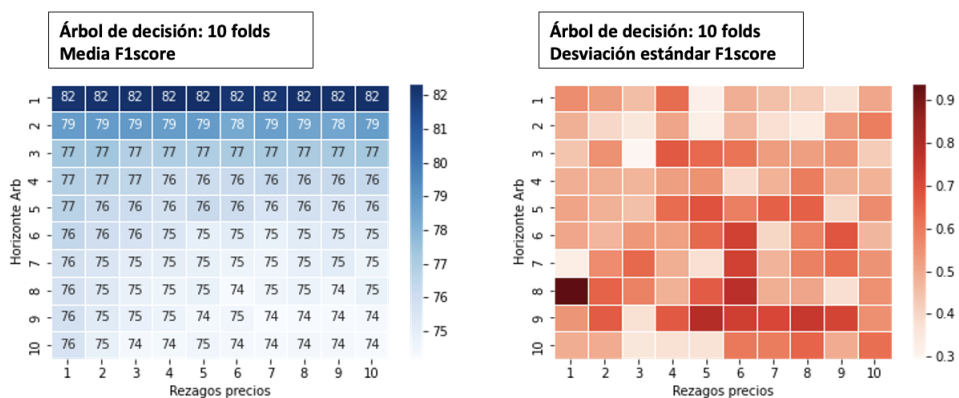


Figura 4: Resultados árboles de decisión

Finalmente en la Figura 5 se encuentran los resultados de clasificación usando máquinas de soporte vectorial. Nuevamente se obtienen *F1-score* más

bajos en la medida que se incrementa el horizonte de arbitraje. En cuanto a la cantidad de información rezagada, se observa el mismo resultado de los árboles de decisión; existe un leve decrecimiento en el *F1-score* a medida que aumenta la cantidad de información rezagada con la que entrenamos el modelo, resulta aún difícil determinar una cantidad de rezagos óptima. Este modelo logra los mejores resultados con *F1-score* de hasta 88% en horizontes de predicción de 1 minuto, así mismo los resultados de *precisión* y *recall* se ubican también sobre 88% para los mismos horizontes de predicción. La desviación estándar de los resultados para todos los casos es inferior a 1%.

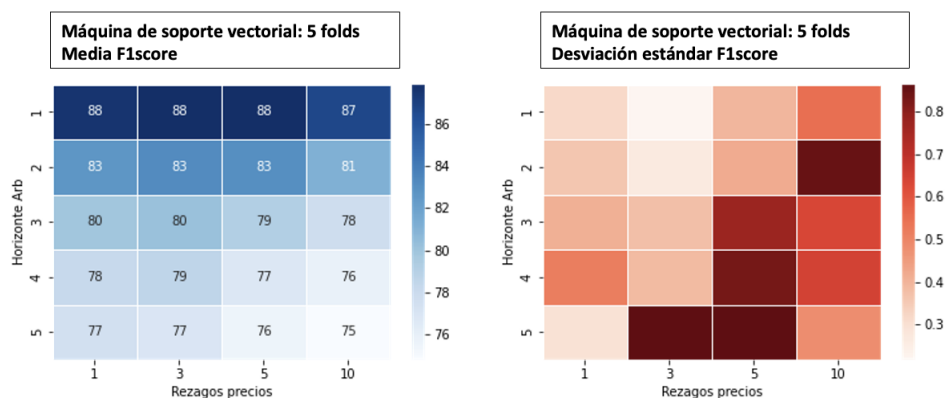


Figura 5: Resultados máquinas de soporte vectorial

6.1.1 Conclusiones exploración

La mejor aproximación para la solución del problema por cuestiones prácticas y de resultados es el de la clasificación multiclase. Todos los modelos muestran mejores resultados en la medida que el horizonte de arbitraje es menor, en ese sentido se obtienen los mejores resultados para predicciones a 1 minuto. Los

resultados para la cantidad de información rezagada que reciben los modelos no muestran ningún patrón claro y deben explorarse más a fondo.

6.2 Exploración II

Para esta segunda fase de la exploración, se utilizan diferentes modelos de Machine Learning y objetivos de clasificación sobre condiciones de arbitraje para diferentes bases de datos de rezagos en los precios. Se realizan algunas observaciones preliminares con base en una primera aproximación a los datos. Se definen varios targets y rezagos para la clasificación así como un modelo base sobre el cual se establecen métricas de desempeño como punto de referencia. Se observa la incidencia de las variables categóricas, la selección de features y el número de lags en las métricas de desempeño. Para esta fase de la exploración se utilizan una progresión de 3, 5 y 10 lags para los rezagos en los precios de las variables continuas y para la variable exógena. Dados los resultados obtenidos en la Exploración I, optamos por predecir si en la observación siguiente (siguiente minuto) el modelo es capaz de predecir con suficiente precisión las condiciones de arbitraje. Se espera que la observación progresiva para diferentes lags ofrezca una idea más clara acerca del número de rezagos óptimo para el modelo. Se utiliza un valor máximo de 10 lags teniendo en cuenta que a medida que se incrementa el número de lags que el modelo debe considerar, aumenta también la carga computacional, por lo cual idealmente, se espera un número de rezagos óptimo en un rango no muy amplio de lags.

6.2.1 Modelo Base de Referencia

Como punto de partida, se eligió un ensemble de mayoría de votos construido a partir de tres instancias de modelos de clasificación simples: una regresión logística, un árbol de decisión y un clasificador de vecinos más cercanos. Se obtuvieron los resultados detallados en la Figura 6 y 7 con una clasificación binaria para un solo tipo de arbitraje con la base de tres rezagos. El modelo trata de predecir si hay o no condiciones de arbitraje para la observación siguiente ³

De este modo, se parte de una precisión cercana al 90% para las observaciones en las que no se presentan condiciones de arbitraje y un recall $> 78\%$ para la predicción de condiciones de arbitraje del tipo BL/SE (Buy local/Sell externo), además se obtiene un recall $> 67\%$ para las condiciones de arbitraje del tipo BE/SL (Buy externo/Sell local).

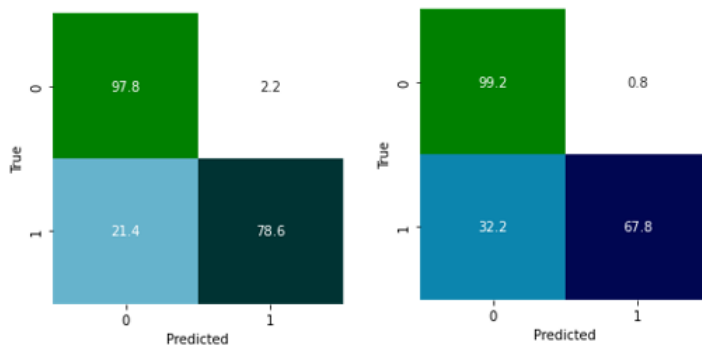


Figura 6: Modelo Base: Clasificación Binaria BL/SE y BE/SL, respectivamente

³Para la matriz de confusión en la izquierda: 0 no arbitraje, 1: Arbitraje BL/SE. Para la matriz de confusión de la derecha: 0 no arbitraje, 1: Arbitraje BE/SL.

```
10-fold cross validation:

Modelo: Logistic regression
Accuracy: 0.91 (+/- 0.00) [Logistic regression]
Precision: 0.89 (+/- 0.01) [Logistic regression]
Recall: 0.68 (+/- 0.01) [Logistic regression]
f1_score: 0.73 (+/- 0.01) [Logistic regression]
Modelo: Decision tree
Accuracy: 0.94 (+/- 0.00) [Decision tree]
Precision: 0.83 (+/- 0.01) [Decision tree]
Recall: 0.92 (+/- 0.01) [Decision tree]
f1_score: 0.87 (+/- 0.01) [Decision tree]
Modelo: KNN
Accuracy: 0.95 (+/- 0.00) [KNN]
Precision: 0.88 (+/- 0.01) [KNN]
Recall: 0.87 (+/- 0.01) [KNN]
f1_score: 0.87 (+/- 0.01) [KNN]
```

Figura 7: Modelo Base: Métricas de desempeño

6.2.2 Otros modelos

Para la base de tres, cinco y diez rezagos, se prueban varios modelos de ensemble con estrategias de bagging, stacking y boosting, incluyendo modelos de última generación de gradient boosting como LigthGBM, Histogram Gradient Boosting y CatBoost ⁴. Para cada categoría por separado y para clasificación multiclase, cada uno de los modelos en mención identifica el tipo de categoría: no arbitraje, BL/SE y BE/SL⁵. De los resultados obtenidos se puede observar en la Figura 8 la matriz de confusión para un modelo XGBoost bastante preciso. En cada caso observamos un aumento sustancial del recall para cada categoría de arbitraje.

De esta forma, se observa una mejora sustancial en el desempeño en términos de accuracy, recall y precisión para todas las categorías de arbitraje aplicando modelos de clasificación multiclase. En particular, para la base de 10 rezagos se evidencia un ligero incremento en las métricas de desempeño. Este resultado apoya la idea de buscar un número de rezagos óptimo para un rango de lags.

⁴Los resultados detallados pueden consultarse en el Anexo 2 del presente documento el cual esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

⁵Clasificación para la matriz de confusión multiclase, 0: no arbitraje. 1: Arbitraje BE/SL. 2: Arbitraje BL/SE

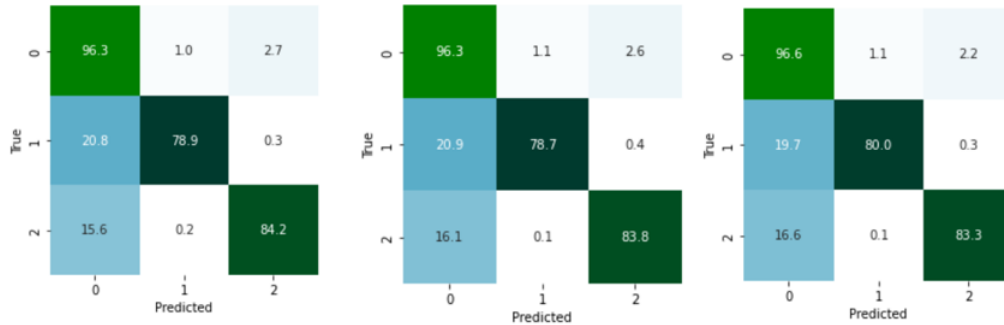


Figura 8: XGBoost: Clasificación Multiclase para 3,5 y 10 lags, respectivamente

	precision	recall	f1-score	support
0	0.96	0.97	0.96	12279
1	0.84	0.80	0.82	944
2	0.85	0.83	0.84	1896
accuracy			0.94	15119
macro avg	0.88	0.87	0.88	15119
weighted avg	0.94	0.94	0.94	15119

Figura 9: Métricas de desempeño para un XGBoost con 10 lags de rezago.

6.2.3 Selección de Features

Se aplica un análisis de selección de componentes principales (PCA) con el propósito de observar si al realizar una reducción de la dimensión, los resultados de los modelos se preservan. A menudo, la reducción de la dimensión disminuye el espacio de tiempo y el almacenamiento requerido, así mismo, la identificación de multicolinealidades entre las variables mejora el rendimiento del aprendizaje modelo.

En la Figura 10 se identifica que las primeras siete componentes principales explican cerca del 90% de la varianza, por lo cual se realiza un pipeline de clasificación con un modelo gradient boosting para las primeras siete componentes

principales cuyos resultados se pueden observar en la Figura 11.⁶:

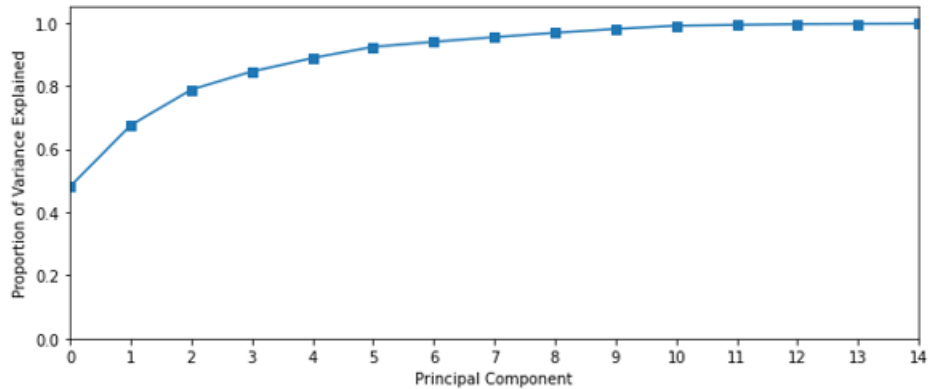


Figura 10: Porcentaje de varianza explicada para 15 componentes principales

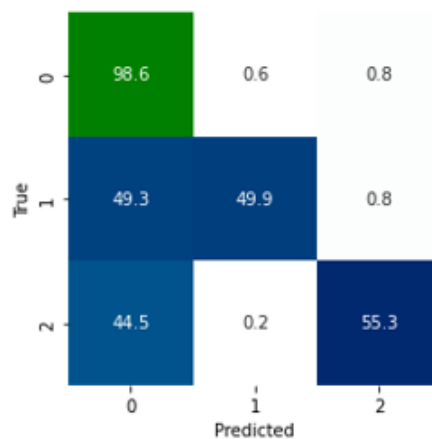


Figura 11: Resultados de la clasificación al aplicar PCA

Se concluye que el modelo pierde capacidad predictiva al aplicar un método de selección de features por PCA (Figura 11). Esto puede obedecer a que la estructura de la serie de tiempo de rezagos se rompe cuando se realiza la selección de features.

⁶Clasificación para la matriz de confusión multiclase, 0: no arbitraje. 1: Arbitraje BE/SL. 2: Arbitraje BL/SE

6.2.4 Variables Categóricas

También se quiere saber si al considerar las variables categóricas, que reflejan las condiciones operativas de los mercados involucrados, los resultados obtenidos se afectan sustancialmente. Para esto, se incluyen las variables ordinales booleanas (festivos, mercados abiertos, subastas, horarios de negociación) dentro de las bases de rezagos. En la Figura 12 se aprecian los resultados para una predicción con un modelo XGBoost para la base de 5 lags:

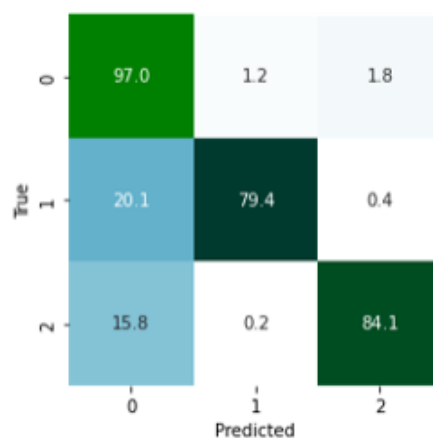


Figura 12: Resultados de la clasificación con variables categóricas

Efectivamente, se evidencia que al incluir las variables ordinales y categóricas en la base las métricas de desempeño presentan un aumento para cada categoría, en particular el recall y la precisión. No obstante, los resultados no son sustancialmente superiores.

6.2.5 Conclusiones Exploración

Las conclusiones extraídas de la metodología de exploración determinaron la construcción de una metodología formal de selección y evaluación de un modelo de

Machine Learning óptimo en la siguiente sección:

- Los modelos de clasificación multiclase ofrecen mejores resultados que los de clasificación binaria. (Tanto en la exploración I como en la exploración II)
- Los modelos de ensemble contruidos a partir de los clasificadores con mejor desempeño ayudan a obtener mejores métricas de desempeño que cada modelo considerado individualmente.
- No es una buena idea aplicar técnicas de reducción de la dimensión para las bases de datos construidas a partir de rezagos.
- A medida que el número de rezagos aumenta, se observaron ligeros incrementos en las métricas de desempeño.
- La inclusión de las características operativas de los mercados como variables categóricas ordinales y booleanas, deriva en una mejora en las métricas de desempeño, no obstante la mejora no es sustancial y el modelo puede trabajar con desempeños similares al considerar solo variables de precio continuas.

6.3 Metodología Formal

Se evalúan los modelos sobre los cuales se tienen las mejores métricas de desempeño en los ejercicios de exploración. Se elige una ventana de rezagos de uno a quince para observar los resultados de evaluar estos modelos para cada uno de los rezagos. Del mismo modo, se observa la capacidad predictiva de los modelos hacia el futuro, es decir; evaluaremos si a partir de las observaciones

actuales es posible tener una buena predicción de las condiciones de arbitraje en observaciones futuras. Se comparan ambos enfoques y se elige un modelo y su ventana de lags óptima con el propósito de realizar una evaluación financiera del desempeño del mismo. Para esto, se define, en términos de la función de arbitraje, una función de backtestig sobre las predicciones y se calcula el estado de pérdidas y ganancias para las predicciones. Se define un grupo de clasificadores. Esta vez se parte de un clasificador base Random Forest para observar si hay un mejor desempeño. Se toma además, los clasificadores con mejores métricas en la evaluación por exploración: una instancia de Gradient Boosting, una instancia de un clasificador LigthGBM y un clasificador Histogram Gradient Boosting. Teniendo en cuenta que el desempeño observado al agregar las variables categóricas a los features no produjo un resultado sustancialmente superior en la evaluación al tanteo, la estimación actual se realizará filtrando este tipo de variables. El efecto de incluir las variables categóricas se observa más adelante en la implementación de un modelo de Deep Learning con capas de embedding.

6.3.1 Evaluación para un rango de lags

Para el rango de lags de uno a quince, se va a evaluar cada uno de los modelos previamente definidos. La función de evaluación, toma los rezagos en los precios de las variables para cada variable de precio y realiza una clasificación multiclase para las condiciones de arbitraje actuales. Para lograr una evaluación más precisa de las métricas de desempeño, se utiliza un cross validation estratificado repetido. En cada repetición se realizan 3 splits para el conjunto de entrenamiento. Se realizan cinco repeticiones y se tabulan los resultados promedio para el accuracy, la precisión y el recall promedio, con sus respectivas desviaciones estándar. Se

produce una matriz de confusión para cada lag y se grafican los resultados para las principales métricas.

Un gráfico de los resultados obtenidos para el modelo Histogram Gradient Boosting se observa en la Figura 13:

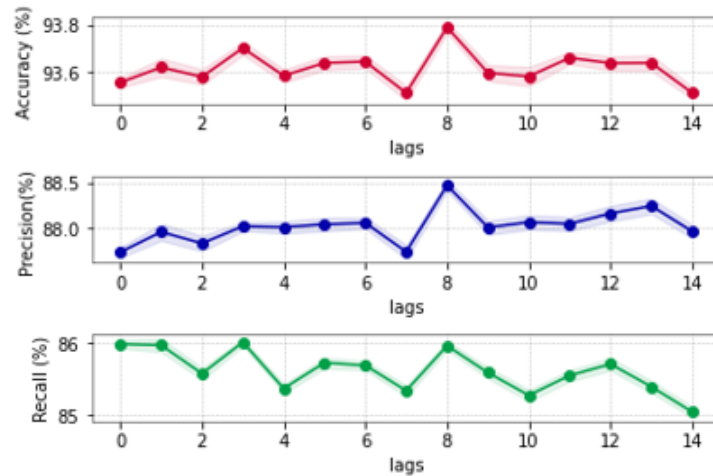


Figura 13: Desempeño para un rango de lags: Histogram Gradient Boosting

Para este caso, observamos el desempeño óptimo para $lags = 8$. Este procedimiento se repite con un modelo por mayoría de votos, construido como un ensemble de los mejores modelos hallados en la fase de exploración. Los resultados tabulados se observan en la Figura 14.

6.3.2 Modelo de mayoría de votos con rango de lags óptimo (MVC)

En la Tabla 1 observamos los resultados para algunas de las métricas de desempeño para el modelo MVC.⁷

⁷Los resultados detallados pueden consultarse en el Anexo 3 del presente documento el cual esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

Modelo	lags	repeticiones	fit time	test accuracy	train accuracy	test precision macro	train precision macro	test recall macro	train recall macro
VotingClassifier	1	15	9,353	0,936	0,949	0,876	0,903	0,863	0,889
VotingClassifier	2	15	13,468	0,938	0,955	0,881	0,916	0,862	0,899
VotingClassifier	3	15	18,768	0,938	0,957	0,881	0,920	0,862	0,906
VotingClassifier	4	15	24,644	0,938	0,959	0,882	0,926	0,860	0,909
VotingClassifier	5	15	30,839	0,937	0,960	0,881	0,929	0,860	0,912
VotingClassifier	6	15	35,731	0,938	0,962	0,884	0,932	0,860	0,915
VotingClassifier	7	15	38,370	0,939	0,964	0,885	0,937	0,862	0,922
VotingClassifier	8	15	46,039	0,939	0,965	0,884	0,938	0,861	0,921
VotingClassifier	9	15	51,926	0,938	0,966	0,883	0,940	0,861	0,926
VotingClassifier	10	15	56,357	0,938	0,965	0,882	0,937	0,862	0,925
VotingClassifier	11	15	60,087	0,939	0,966	0,884	0,939	0,863	0,925
VotingClassifier	12	15	535,206	0,938	0,966	0,882	0,938	0,862	0,925
VotingClassifier	13	15	62,866	0,940	0,969	0,885	0,945	0,863	0,933
VotingClassifier	14	15	64,676	0,938	0,969	0,881	0,944	0,859	0,933
VotingClassifier	15	15	68,590	0,939	0,968	0,883	0,945	0,862	0,931

Tabla 1: Resultados métricas para un rango de lags: MVC

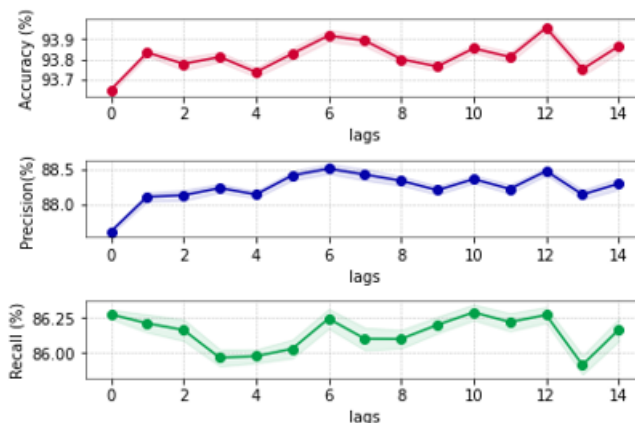


Figura 14: Resultados métricas para un rango de lags: MVC

Se identifica que todas las métricas de desempeño, incluyendo el recall promedio, muestran mejores resultados comparados con la evaluación de cada uno de los modelos por separado. Así mismo, la precisión para cada una de las categorías de arbitraje presenta una leve mejora. Como se observa en la Figura 15, los resultados óptimos se obtienen para $lags = 12$. Caso en el que se observa un accuracy cercano a 94% y un $F1$ -score aproximado de 87%. En la siguiente sección, se analiza el efecto de predicción a futuro para las condiciones de arbitraje. Si los resultados no son mejores, se tomará el modelo de clasificación por mayoría de votos como modelo óptimo para la evaluación financiera.

6.3.3 Modelos de evaluación a Futuro

Es de esperar que el modelo pierda capacidad predictiva hacia al futuro. Para confirmarlo, se realiza un análisis de autocorrelación de las observaciones de las variables de spread de precios:

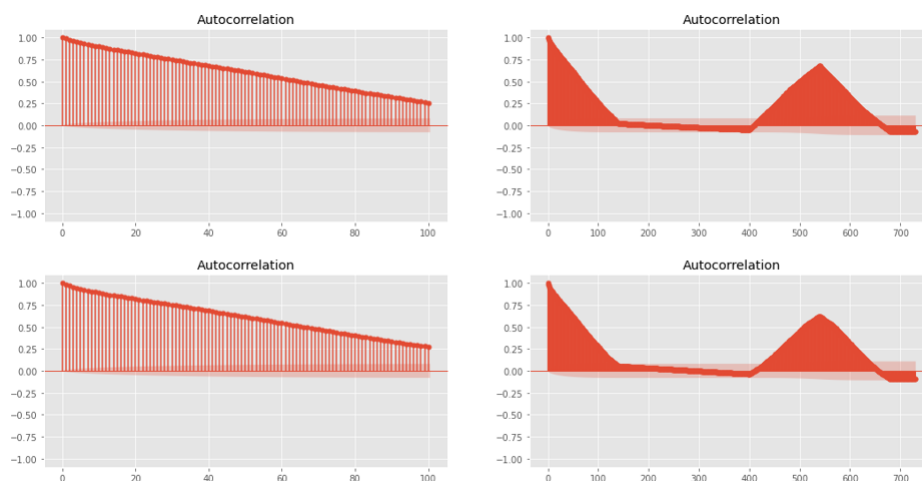


Figura 15: Autocorrelación de corto (izq.) y largo(der.) plazo para los Spreads

Tanto las figuras de autocorrelación como las autocorrelaciones parciales indican que el modelo pierde memoria a medida que se incrementa el horizonte de tiempo. En el caso de la autocorrelación parcial, la Figura 16 indica que la memoria de la serie de spreads está entre una y dos observaciones⁸. Así mismo, los resultados que se obtienen al aplicar los modelos hacia un horizonte futuro en un rango de lags de 1 a 15 refuerzan esta conclusión (17):

De este modo concluimos que el enfoque óptimo es el de considerar rezagos en los precios y una predicción a futuro solo para la observación siguiente.⁹

⁸Sobre la metodología Box Jenkins para análisis de autocorrelación en series de tiempo, se puede consultar [26]

⁹Los resultados detallados pueden consultarse en el Anexo 3 del presente documento el cual

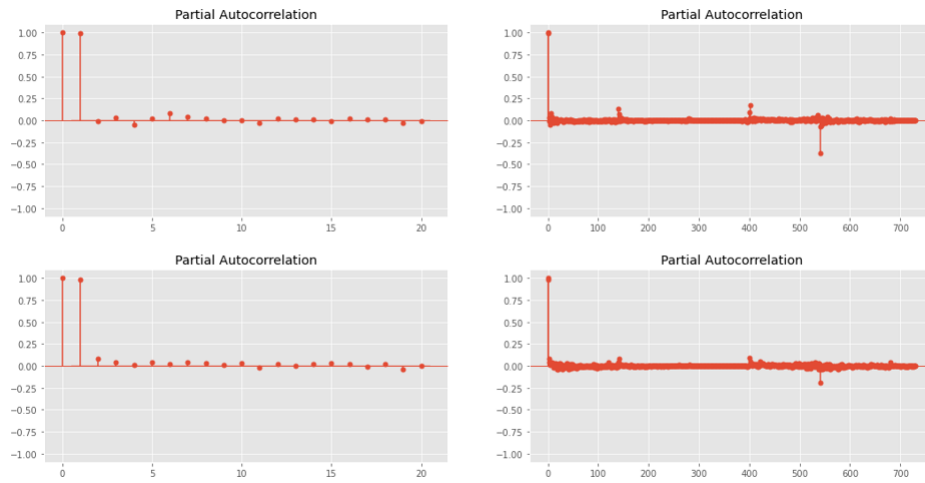


Figura 16: Autocorrelación parcial de corto (izq.) y largo (der.) plazo para los Spreads

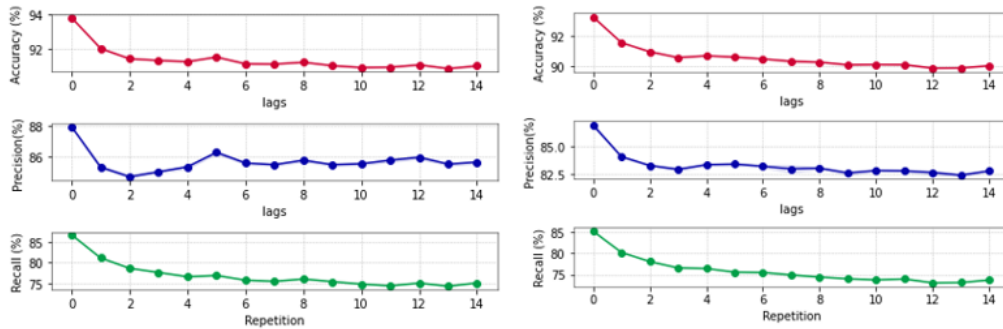


Figura 17: Resultados evaluación horizonte de tiempo futuro de 1 a 15 lags: Histotgam Gradient Boosting y LightBGB

6.4 Modelos de Deep Learning

Se construyen y evalúan algunos modelos de Deep Learning con el objetivo de analizar su rendimiento en términos de clasificación de las condiciones de arbitraje para la base de datos de lags de precios. Al final, se pretende comparar los esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

resultados contra los obtenidos con el modelo de técnicas de Machine Learning, por lo que se utiliza una base con 12 retrasos en los precios.

6.4.1 Arquitecturas

Se utilizan tres arquitecturas distintas: Para la primera, el modelo utiliza capas de embedding con el objetivo de transformar las variables categóricas y ordinales en vectores continuos que puedan explicar mejor la incidencia de estas variables en las condiciones de arbitraje mediante relaciones de similaridad, las capas de embedding para las variables categóricas se concatenan con la capa continua y al final se conecta con una capa densa con función de activación softmax de tres unidades para observar si se logra una clasificación adecuada de las condiciones de arbitraje. El entrenamiento de las capas de embedding se realiza al mismo tiempo que se entrena el modelo. Para la segunda arquitectura, se utilizarán varias capas densas fully connected y para la tercera arquitectura, se utilizarán capas convolucionales y de max pooling con strides cortos, teniendo en cuenta lo observado en la sección anterior acerca de la memoria del modelo. En ambos casos, la función de activación final será una función softmax de tres unidades y se entrenarán los modelos con las variables numéricas.

6.4.2 Resultados arquitectura Embeddings

Para el entrenamiento se utilizaron conjuntos de entrenamiento y validación independientes, 250 epochs de entrenamiento, un batch size de 200 y un optimizador del gradiente Adam con tasa de aprendizaje de 0.00001. Los resultados del entrenamiento y la evaluación sobre un conjunto de prueba que no se utilizó en el entrenamiento se muestran en la figuras 18 y 19.

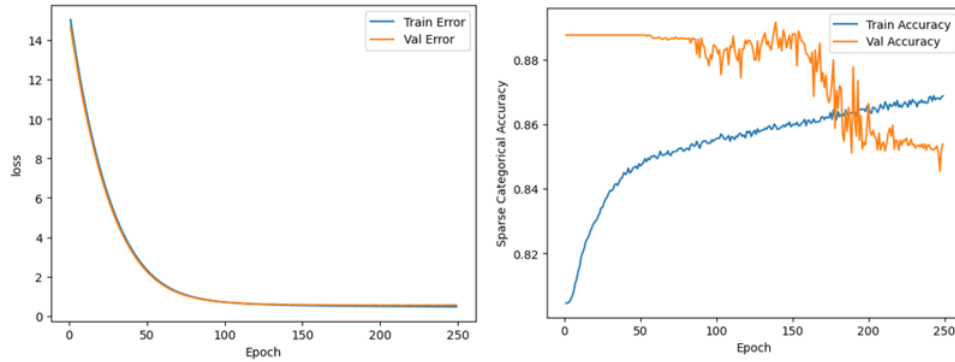


Figura 18: Resultados entrenamiento arquitectura de embeddings.

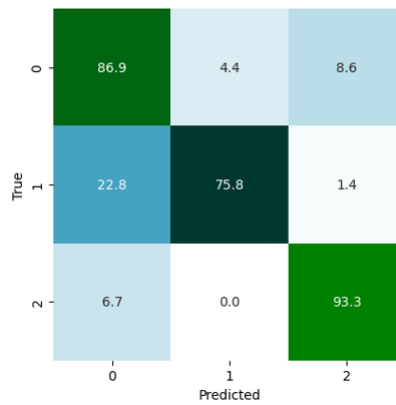


Figura 19: Resultados predicción con embeddings (test set independiente.)

6.4.3 Resultados Arquitectura capas densas

Para el entrenamiento con el modelo de capas densas se utiliza una función de optimización Adam para el descenso en el gradiente con tasa de aprendizaje 0.00001, un número de epochs de 1000 y un batch size de 200. Se observa un menor sobreajuste para el entrenamiento:

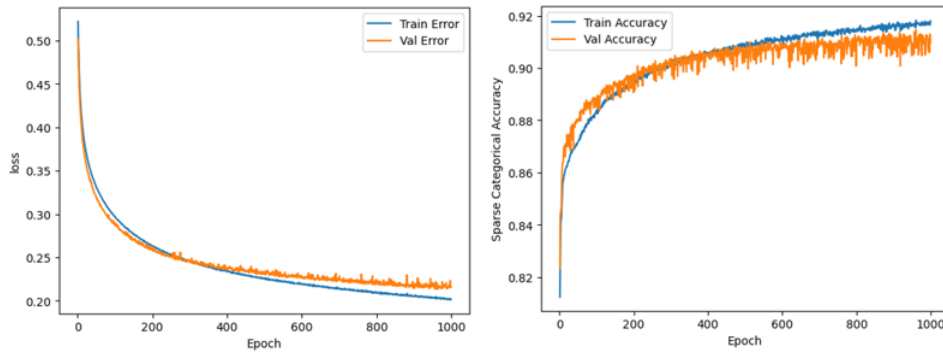


Figura 20: Resultados entrenamiento arquitectura de capas densas.

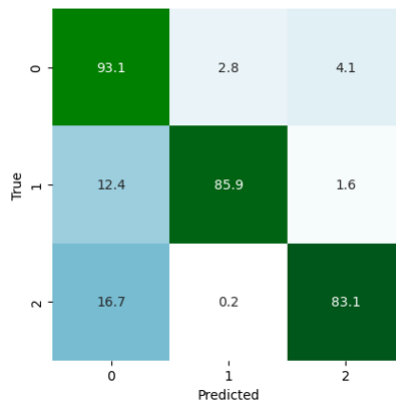


Figura 21: Resultados predicción con capas densas (test set independiente.)

6.4.4 Resultados arquitectura convolucional

Para el entrenamiento con el modelo de capas convolucionales utilizamos una función de optimización Adam para el descenso en el gradiente con tasa de aprendizaje 0.00001, un número de epochs de 700 y un batch size de 100. En este caso también se observa un menor sobre ajuste para el entrenamiento:

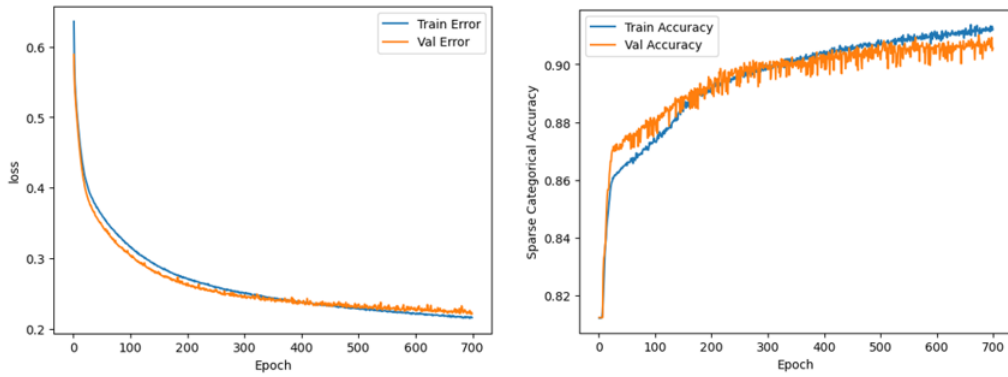


Figura 22: Resultados predicción con capas convolucionales (test set independiente)

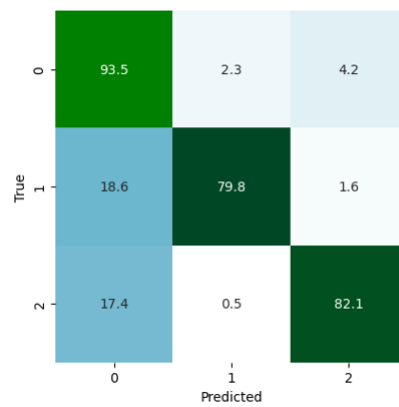


Figura 23: Resultados predicción con capas convolucionales (test set independiente)

6.4.5 Métricas de desempeño Deep Learning

En la Figura 24 se evidencia un aumento en la precisión para las clases de arbitraje. No obstante una pérdida en el recall. Esto refleja el sobreajuste en el entrenamiento de los modelos.

Classification Report (Embedding / Capas Densas / Convolucional)

	precision	recall	f1-score	support
0	0.8693	0.9880	0.9249	12279
1	0.7576	0.3443	0.4734	944
2	0.9332	0.3613	0.5209	1896
accuracy			0.8692	15119
macro avg	0.8534	0.5645	0.6397	15119
weighted avg	0.8703	0.8692	0.8460	15119

	precision	recall	f1-score	support
0.0	0.93	0.97	0.95	12279
1.0	0.86	0.62	0.72	944
2.0	0.83	0.72	0.77	1896
accuracy			0.92	15119
macro avg	0.87	0.77	0.81	15119
weighted avg	0.91	0.92	0.91	15119

	precision	recall	f1-score	support
0.0	0.93	0.96	0.95	12279
1.0	0.80	0.68	0.73	944
2.0	0.82	0.71	0.76	1896
accuracy			0.92	15119
macro avg	0.85	0.79	0.82	15119
weighted avg	0.91	0.92	0.91	15119

Figura 24: Métricas de desempeño para los modelos de Deep learning. Embedding(superior), capas densas(medio), convolucional(inferior)

6.5 Evaluación Financiera

Una vez hecho el barrido por los modelos de Machine y Deep Learning y comparar los resultados, se opta por el modelo de ensemble por mayoría de votos con una base de $lags = 12$ para realizar una evaluación práctica. Se presenta un análisis de los resultados obtenidos por la predicción con el clasificador de mayoría de votos, con el fin de establecer, con mayor precisión, los resultados financieros esperados para la predicción en condiciones operativas de trading

usuales. Para la predicción, se asume que cada vez que el modelo predijo una condición de arbitraje, bien sea de la forma (BE/SL) o (BL/SE), se ejecutó una operación de arbitraje con un tamaño equivalente a 10 mil acciones.

6.5.1 Filtro periodo exdividendo

Durante el periodo de muestra, la acción preferencial de Bancolombia y su respectivo ADR presentaron un periodo ex dividendo (periodo en el que la compañía paga dividendos). El periodo ex dividendo para ambos títulos no es el mismo y las predicciones para las condiciones de arbitraje y de spread se ven afectadas por este tipo de situaciones. Esto ocasiona spreads que reflejan las diferencias entre las fechas de inicio, duración y pago del periodo ex-dividendo para cada instrumento, por lo cual para el análisis financiero de los resultados para las condiciones de arbitraje se pueden considerar dos opciones: incluir o eliminar los dividendos del spread observado (sumar o restar el dividendo del spread de la predicción, según sea el caso), o simplemente; eliminar las observaciones del periodo en el cual ambos títulos estuvieron en periodo ex-dividendo. Para el análisis de los resultados, lo más conveniente es optar por esta última acción, teniendo en cuenta que el spread observado por dividendos no necesariamente genera utilidades en términos de arbitraje a pesar de que el spread ocasionado por dividendos sí origina spreads de precio entre ambos instrumentos.

6.5.2 Tratamiento días festivos

Teniendo en cuenta que al observar spreads durante una fecha festivo, uno de los mercados está cerrado y en la práctica no es posible ejecutar operaciones en todo momento si la predicción así lo indica, el análisis financiero de los resultados en

festivos debe hacerse de forma diferencial. Inicialmente, se filtran los festivos de la base y se obtienen los resultados solamente de las predicciones para sesiones de mercado abierto y subastas. Esto reflejará un resultado más preciso y ajustado a las posibilidades prácticas.

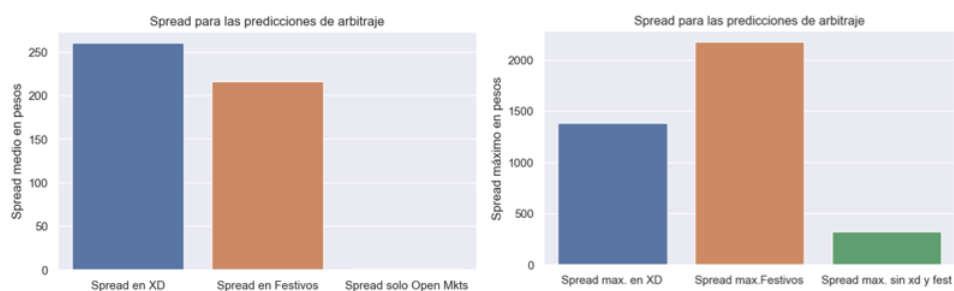


Figura 25: Efecto xd y festivos en el spread de las predicciones.

6.5.3 Resultados sesión abierta con subastas

Al eliminar los festivos y las fechas ex dividendo, se logra obtener un análisis preliminar del resultado de los trades que, en la práctica, reflejan las condiciones más ajustadas a la realidad operativa cuando ambos mercados están abiertos. Los resultados generales, incluyendo pérdidas, ganancias y las predicciones en las cuales se predijo una condición de no arbitraje se aprecian en la *Tabla 2* y *Figura 26*¹⁰.

¹⁰Los resultados detallados pueden consultarse en el Anexo 2 del presente documento el cual esta alojado en el repositorio de GitHub: [link a github cambiar!!](#)

	Resumen Trades (COP)	Resumen Spread (COP)
cantidad	13034	13034
media	\$ 13.453,69	1.345
std	\$ 185.235,67	18,524
min	-\$ 2.531.065,00	-253,1
max	\$ 3.208.802,55	320,88

Tabla 2: Descriptivo de los resultados para las observaciones y el spread de las predicciones.

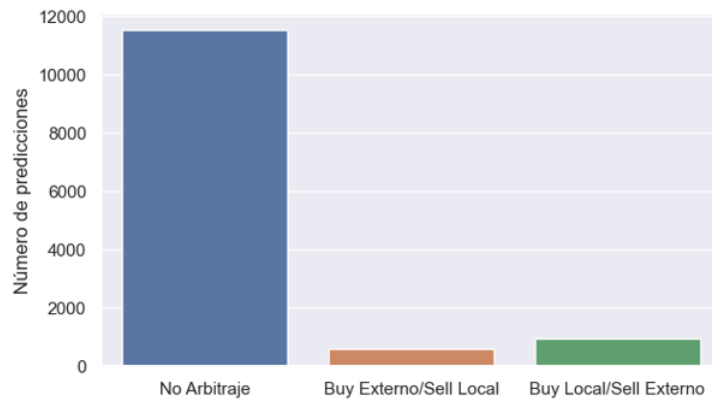


Figura 26: Número de predicciones para cada categoría de arbitraje.

6.5.4 Pérdidas y Ganancias

Para determinar si el modelo resulta rentable, se observa el número de trades positivos y negativos y los valores medios de las pérdidas y las ganancias. Idealmente, debería ocurrir que el número de trades positivos exceda al número de trades negativos y que el valor medio de las ganancias sea superior al valor absoluto de la media de las pérdidas, tal como se evidencia a continuación (Figura 27 y *Tabla 3*):



Figura 27: Número de trades positivos y negativos.

	Trades Positivos (COP)	Trades Negativos(COP)
cantidad	928	594
media	\$ 432.499,44	-\$ 380.478,10
std	\$ 357.328,99	\$ 347.504,75
min	\$ 949,30	-\$ 2.531.065,00
25%	\$ 209.190,12	-\$ 535.450,00
50%	\$ 372.033,95	-\$ 296.820,00
75%	\$ 564.563,12	-\$ 123.130,75
max	\$ 3.208.802,55	-\$ 343,75

Tabla 3: Descripción para trades positivos y negativos.

6.5.5 Pérdidas y ganancias incluyendo días festivos

Teniendo en cuenta que en los días festivos, alguno de los mercados permanece cerrado, no es posible ejecutar operaciones. No obstante, en la definición de la función de arbitraje, asumimos que en condiciones de días festivos hay arbitraje cuando se observa spread entre el precio de cierre del último día hábil antes del festivo y el precio en el mercado dual que permanece abierto. En este caso, es posible realizar operaciones en el mercado dual que permanece abierto, pero la operación contraria sólo se puede ejecutar al día siguiente en la apertura del nuevo

día hábil. Esta condición fue incluida en la función de evaluación financiera. De esta manera, es posible que, durante un día festivo, la predicción indique que se pueden realizar operaciones de arbitraje en todo momento, pero en la práctica esto no es posible de ejecutar. Lo usual en la práctica del arbitrador es que, al observar un spread de precio entre los precios de cierre en el mercado cerrado y el precio que se actualiza constantemente en el mercado abierto, se ejecuta una sola operación en el mercado abierto (usualmente al cierre del mercado) y la operación contraria se ejecuta en la apertura del siguiente día hábil en el mercado cerrado por el festivo. A continuación se incluye esta condición operativa en el análisis de resultados y se observa el efecto de incorporarla en la estrategia de trading, según las predicciones del modelo (Figura 28).

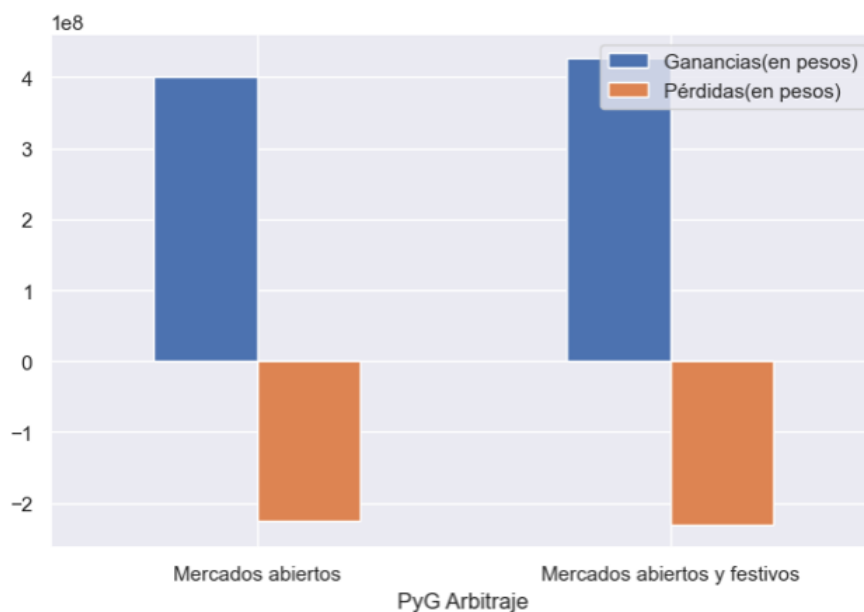


Figura 28: Pérdidas y las ganancias en COP con y sin días festivos.

Se observa en la *Tabla 4* que al incluir una operación por cada festivo en condiciones de arbitraje, la utilidad neta de la estrategia puede aumentar cerca del 10% .

	Mercado Abierto	Festivos	Total
Neto Trades Positivos	\$ 401.359.479,90	\$ 26.402.539,50	\$ 427.762.019,40
Neto Trades Negativos	-\$ 226.003.993,10	-\$ 4.084.368,00	-\$ 230.088.361,10
Total	\$ 175.355.486,80	\$ 22.318.171,50	\$ 197.673.658,30

Tabla 4: Estado de pérdidas y ganancias neto para el conjunto de prueba.

7 CONCLUSIONES

De este modo se concluye que es posible construir un modelo rentable de arbitraje algorítmico que se ajusta a las características operativas de los mercados involucrados utilizando técnicas de Machine y Deep Learning. Los resultados observados para distintos enfoques son muy similares; para los modelos de Deep Learning es muy importante lograr un entrenamiento preciso y un adecuado ajuste de los hiperparámetros. Para los mejores modelos, se obtuvieron métricas de accuracy superiores al 90% y precisiones superiores al 85% para predicciones de clasificación de arbitraje multiclase de corto plazo. Para el recall se observan mayores desviaciones estándar, sobre todo en los modelos base y de deep learning, no obstante, para este último caso, concluimos que las desviaciones se pueden corregir con entrenamientos más ajustados. Los modelos de ensemble para clasificación multiclase ofrecen los mejores resultados y actúan como un potenciador para los resultados con modelos de clasificación más simples. Los modelos por mayoría de votos logran mejores clasificaciones. Concluimos que la capacidad predictiva de los modelos, como es de esperar, disminuye con el tiempo. La mejor

predicción se obtiene para la observación siguiente. Sin embargo, el decaimiento de la memoria no implica que no se puedan obtener modelos rentables de operación para horizontes de tiempo superiores. En términos de la información, los mejores modelos se construyeron a partir de bases de datos que contienen información rezagada en los precios. En general, considerar un mayor número de observaciones rezagadas en los precios deriva en un aumento en las métricas de desempeño. En nuestro caso particular obtuvimos que para nuestro modelo de predicción de corto plazo un número de lags óptimo está al rededor de 12 rezagos, sin embargo, para cada activo, cada horizonte de tiempo y cada modelo, el número de rezagos óptimo actúa como un hiperparámetro del modelo que es mejor tunear con cuidado. Es muy importante considerar las características operativas de los mercados involucrados. La inclusión detallada de las características operativas a modo de variables categóricas no solo implica mejores métricas de desempeño sino también una mejor evaluación del desempeño en la práctica. Los resultados financieros arrojan resultados positivos en términos de pérdidas y ganancias: el modelo construido genera retornos dentro del umbral aceptable para las instituciones financieras que realizan este tipo de operaciones. El enfoque de Machine Learning funciona adecuadamente para el mercado de renta variable colombiano. Las técnicas consideradas se presentan como una alternativa para el desarrollo de nuevos tipos de operación en el mercado de capitales colombiano y por lo tanto, un camino de desarrollo y crecimiento para el mismo en el futuro cercano.

Referencias

- [1] Dominik Rösch. The impact of arbitrage on market liquidity. *Journal of Financial Economics*, 142, 05 2021.
- [2] Jim Fischer. Modern portfolio theory and the efficient markets hypothesis: How well did they serve canada’s baby-boom generation? 10 2019.
- [3] Eugene Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.
- [4] J. M. Samuels. Inefficient capital markets and their implications. 1981.
- [5] Shireen Tang, Shijie Huang, Elizabeth Bowman, Carsten Murawski, and Peter Bossaerts. The efficient markets hypothesis does not hold when securities valuation is computationally hard. *SSRN Electronic Journal*, 01 2017.
- [6] Jonathan B. Berk. A critique of the efficient market hypothesis. 2008.
- [7] Veliota Drakopoulou. A review of fundamental and technical stock analysis techniques. *Journal of Stock Forex Trading*, 05, 01 2016.
- [8] Mohamed Masry. The impact of technical analysis on stock returns in an emerging capital markets (ecm’s) country: Theoretical and empirical study. *International Journal of Economics and Finance*, 9:91, 02 2017.
- [9] Sophie Emerson, Ruairí Kennedy, Luke O’Shea, and John R. O’Brien. Trends and applications of machine learning in quantitative finance. *Machine Learning eJournal*, 2019.

- [10] Benoit B. Mandelbrot. *The variation of certain speculative prices*, pages 371–418. Springer New York, New York, NY, 1997.
- [11] S. Patterson. *Dark Pools: The Rise of the Machine Traders and the Rigging of the U.S. Stock Market*. Crown, 2013.
- [12] S. Patterson. *The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed It*. Crown, 2010.
- [13] Zhe Huang and Franck Martin. Optimal pairs trading strategies in a cointegration framework. July 2017. working paper or preprint.
- [14] Sayad Baronyan, I. Ilkay Boduroglu, and EMRAH ŞENER. Investigation of stochastic pairs trading strategies under different volatility regimes. *The Manchester School*, 78:114 – 134, 09 2010.
- [15] K. J. Hong and S. Satchell. Time series momentum trading strategy and autocorrelation amplification. *Quantitative Finance*, 15(9):1471–1487, 2015.
- [16] Andy Lin and Illya Barziy. Pairs trading based on renko and kagi models, Oct 2021.
- [17] John Moody and Matthew Saffell. Learning to trade via direct reinforcement. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 12:875–89, 07 2001.
- [18] Riccardo Rosati, Luca Romeo, Carlos Alfaro Goday, Tullio Menga, and Emanuele Frontoni. Machine learning in capital markets: Decision support system for outcome analysis. *IEEE Access*, 8:109080–109091, 2020.

- [19] James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick. Reinforcement Learning and Savings Behavior. *Journal of Finance*, 64(6):2515–2534, December 2009.
- [20] G Andrew Karolyi and Stijn Van Nieuwerburgh. New Methods for the Cross-Section of Returns. *The Review of Financial Studies*, 33(5):1879–1890, 02 2020.
- [21] Paul Yoo, Maria H. Kim, and T. Jan. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 2:835–841, 2005.
- [22] Patrick Jaquart, David Dann, and Christof Weinhardt. Short-term bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, 7:45–66, 2021.
- [23] Adriano Koshiyama, Nick Firoozye, and Philip Treleaven. Algorithms in future capital markets: A survey on ai, ml and associated algorithms in capital markets. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2021. Association for Computing Machinery.
- [24] Nicolas Huck. Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, 278(1):330–342, 2019.

- [25] Baoqiang Zhan, Shu Zhang, Helen Du, and Xiaoguang Yang. Exploring statistical arbitrage opportunities using machine learning strategy. *Computational Economics*, 60, 11 2021.
- [26] George EP Box and Gwilym M Jenkins. Some comments on a paper by chatfield and prothero and on a review by kendall. *Journal of the Royal Statistical Society. Series A (General)*, 136(3):337–352, 1973.