



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

ANÁLISIS DE PREDICTIBILIDAD EN ACTIVOS FINANCIEROS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

ALEJANDRO ANTONIO MORALES PEZOA

PROFESOR GUÍA:
NÉSTOR BECERRA YOMA

MIEMBROS DE LA COMISIÓN:
DAVID DÍAZ SOLÍS
ALEJANDRO BERNALES SILVA
ÁLVARO SILVA MADRID

SANTIAGO DE CHILE
2023

ANÁLISIS DE PREDICTIBILIDAD EN ACTIVOS FINANCIEROS

La presente investigación se centra en la aplicación del método denominado Linear Predictive Coding (LPC) en el ámbito del mercado financiero, específicamente a la predicción de precios de cierre de las acciones pertenecientes al índice S&P500. El objetivo principal de este estudio es explorar la predictibilidad de ciertos activos financieros utilizando este método, además de examinar su viabilidad y eficacia como herramienta predictiva dentro del contexto financiero.

Se llevó a cabo una adaptación del método LPC, originalmente empleado para la estimación de señales de voz, con la finalidad de aplicarlo en la predicción de los precios de las acciones. Por otro lado, se recopilan datos históricos que abarcan un período de 12 años desde Agosto el 2010 hasta Agosto del 2022 y en estos se incluyen los precios diarios de cierre de las acciones S&P500 que pertenecen al mercado estadounidense.

Una vez obtenidas las predicciones, estas se proceden a evaluar en tres diferentes métricas: Mean Absolute Error (MAE), Directional Accuracy (DA) métricas ampliamente utilizadas en la literatura y se propone una nueva métrica llamada Discriminante (DIS), la cual consta en la relación de las autocorrelaciones entre la señal original y la predicha. Se otorgó mayor énfasis a la métrica DA, dado su comportamiento y relevancia, donde al predecir la dirección de los activos podría generar un mayor impacto a la hora de invertir dinero.

Posteriormente, se realizaron simulaciones de trading, en base a las predicciones obtenidas, donde se observó que; aunque la métrica DA contribuye en términos generales a obtener mayores retornos, no es lo único que se debe considerar para maximizar las ganancias monetarias, debido a que se identificaron otros factores que pueden ser determinantes a la hora de invertir, como lo es su comportamiento a través del tiempo, es decir, su nivel de riesgo.

La comparación con el modelo predictivo ARIMA (Autoregressive Integrated Moving Average) permitió demostrar la robustez del modelo LPC, debido a los similares resultados en las métricas, demostrando que mejoró MAE y DIS en comparación a LPC, pero no logró superarlo en DA, demostrando que se logró competir con ARIMA, a pesar de que este último sea una técnica mas compleja y mucho mas utilizada en este campo.

En cuanto a robustez y confiabilidad de los resultados, cabe destacar que existe la posibilidad de replicar la técnica empleada en otras acciones, mercados y períodos de tiempo. Si bien no se llevaron a cabo nuevos experimentos en otros mercados, los hallazgos obtenidos indican que podrían aplicarse de forma similar en otros contextos. Asimismo, esta posibilidad se basa en la solidez de los resultados y el enfoque metodológico de la investigación.

Asimismo, los resultados y sus respectivas conclusiones en esta investigación proporcionan valiosa información sobre la predictibilidad de las acciones del índice S&P500 utilizando el método LPC. Al identificar los factores determinantes que influyen en la predictibilidad de las acciones, se logró obtener una comprensión más profunda del comportamiento de mercado.

*Se lo dedico a todos los que me vieron crecer y me tendieron una mano.
Pero por sobre todo a mi tío Leonardo que a pesar de ya no estar aquí,
siempre ha sido una gran inspiración para mí.*

Saludos

Agradecimientos

Primero quiero agradecer a mis padres, Patricia y Patricio, quienes me brindaron todo el apoyo que necesitaba desde que nací hasta el día de hoy, siendo pilares fundamentales en mi vida, dándome todas las herramientas necesarias para estar en el punto donde estoy ahora. Siempre dando recomendaciones, consejos y ayudas, pero de igual manera respetando mis decisiones, promoviendo que también creciera como persona.

A mis hermanas mayores, Cristina y Karina, las cuales desde pequeño fueron un ejemplo para mí. Me inspiraron a estudiar siempre con una competitividad sana, mientras veía como ellas pasaban este tipo de procesos en sus respectivas carreras, me hicieron sentir motivado y también admirarlas por lo que hicieron.

Agradezco a todos los profesores y auxiliares de la universidad, sobre todo a aquellos que adicionalmente al conocimiento que me aportaron, fueron más allá, dando consejos, ayudando o simplemente se tomaron el tiempo para explicar nuevamente lo que estaban exponiendo. Gracias también a mis profesores guía y co-guía, Néstor Becerra y David Díaz, los cuales confiaron en mí para trabajar en esta investigación y me guiaron durante un año para poder completar este trabajo, que pese a ser un área inexplorada para mí en ese entonces, terminé siendo de mi agrado, reuniéndonos semanalmente hasta lograr la finalización de mi trabajo de título. También a mi compañero Andre con quien empecé este camino de la investigación y siempre que pudo me tendió una mano.

Un gran soporte fueron mis amigos, a los cuales quiero agradecer por hacer la vida universitaria más agradable, siempre ofreciendo una mano cuando alguien no estaba bien fuese en el ámbito personal o universitario, enseñando, divirtiéndonos, etc. También a mis amigos del colegio, los cuales fueron un gran apoyo durante la carrera, pero sobre todo en pandemia con diversas reuniones online y videojuegos. Que me ayudaron a despejar la mente y tener una mayor tranquilidad.

Finalmente dar las gracias a mi novia, Valentina, la cual este último año ha sido de un gran apoyo para mí, dándome una motivación extra que antes no tenía, teniendo conversaciones larguísimas muy enriquecedoras, ayudándome a escapar del estrés de la universidad y a la vez orientándome para seguir trabajando, siendo una bonita relación apoyándonos el uno al otro.

Tabla de Contenido

1. Introducción	1
2. Antecedentes en la Predicción de Activos Bursátiles	4
3. Predicción Lineal Aplicada a Activos Financieros	15
3.1. LPC	15
3.1.1. Teorema de Descomposición de Wold	16
3.1.2. Modelo Autorregresivo	16
3.1.3. Análisis LPC	16
3.1.4. Método de Ventanas Deslizantes	17
3.2. Predictibilidad	18
3.2.1. Métricas	18
3.3. Simulación de Trading	19
3.3.1. Métricas	19
3.4. Determinantes de Predictibilidad	19
3.4.1. CUSUM	19
3.4.2. Variables Económicas	20
3.4.3. Regresión Lineal Múltiple	22
3.4.3.1. Condiciones	22
3.4.3.2. Pruebas Estadísticas	23
4. Metodología	25
4.1. Metodología para la implementación de LPC	25
4.1.1. Limpieza de datos	25
4.1.2. LPC	26
4.2. Metodología para la Predictibilidad	27
4.3. Metodología para la simulación de trading	27
4.3.1. Estrategias	27
4.4. Metodología para el Análisis de Determinantes	29
4.4.1. Tipos de Variables	29
4.4.2. Selección de Variables	30
5. Experimentos y Resultados	32
5.1. LPC, Modelamiento, Optimización y Robustez	32
5.1.1. Optimización: Búsqueda de hiperparámetros	33
5.1.2. Robustez	35
5.2. Predictibilidad	35
5.2.1. Optimización de Métricas	35

5.2.2.	Representación de señales	40
5.2.3.	Relación entre Métricas	42
5.3.	Predictibilidad ARIMA	46
5.4.	Simulación de trading	48
5.4.1.	Optimizando DA	48
5.5.	Determinantes de Predictibilidad	51
5.5.1.	Análisis de Quiebres	51
5.5.2.	Análisis de Tendencias	52
5.5.3.	Análisis de Volumen	56
5.5.4.	Regresiones Multivariantes	58
5.5.5.	SET 1	58
5.5.6.	SET 2	59
5.5.7.	SET 3	59
6.	Conclusión	61
	Bibliografía	63
	Anexos	67
	Anexo A. Simulación de Trading para MAE	67
A.1.	Optimizando MAE	67
	Anexo B. Simulación de Trading para DIS	69
B.1.	Optimizando DIS	69
	Anexo C. Cusum	71
	Anexo D. Simulación de trading	73
D.1.	Regresiones para el set 1	73
D.2.	Regresiones para el set 2	74
D.3.	Regresiones para el set 3	75

Índice de Tablas

2.1.	Publicaciones seleccionadas capacidad predictiva más elevada reportadas en la literatura.	5
2.2.	Tipos de variables y su literatura correspondiente	13
4.1.	Variables utilizadas para el análisis de determinantes	30
5.1.	Top 5 acciones optimizadas para DA, MAE y DIS	44
5.2.	Los 10 mejores valores de DA para LPC	46
5.3.	Resultados en ARIMA para las 10 mejores acciones	46
5.4.	Las 10 acciones con mas bajo DA	47
5.5.	10 peores acciones en DA, corridas con ARIMA	48
5.6.	Top 10 acciones del Set 1	48
5.7.	Top 10 mejores acciones para el set 2	49
5.8.	Las 10 acciones con mejores retornos para el Set 3 optimizando DA.	49
5.9.	Estrategias en acciones que presentan mayor retorno sobre buy and hold	49
5.10.	Regresión lineal múltiple para explicar el DA en el set 1	58
5.11.	Resultados del modelo para el set 1	59
5.12.	Regresión lineal múltiple para explicar el DA en el set 2	59
5.13.	Resultados del modelo para el set 2	59
5.14.	Regresión lineal múltiple para explicar el DA en el set 3	59
5.15.	Resultados del modelo para el set 3	60
A.1.	Top 10 acciones para el set 1	67
A.2.	Top 10 acciones para el set 2	67
A.3.	Top 10 acciones para el set 3	68
A.4.	Estrategias en acciones que presentan mayor retorno sobre buy and hold	68
B.1.	Top 10 acciones para el set 1	69
B.2.	Top 10 acciones para el set 2	69
B.3.	Top 10 acciones del set 3	70
B.4.	Estrategias en acciones que presentan mayor retorno sobre buy and hold	70
D.1.	Set 1 Determinantes para el DA	73
D.2.	Set 2 determinantes para el DA	74
D.3.	Set 3 Determinantes para el DA	75

Índice de Ilustraciones

2.1.	Clasificación de Información - Data Collection. Fuente: Zhang, Fan, Ye Ding, and Yuhao Liao (2022).	6
2.2.	Distribución de las combinaciones de las fuentes de datos. Fuente: Jiang, W. (2021).	7
2.3.	Granularidad de los datos utilizados en cada estudio. Fuente: Fuente: Kumbure, Lohrmann, Luukka, and Porras (2022).	8
2.4.	Diversidad de métodos utilizados.	9
2.5.	Publicaciones de predicciones de acuerdo a la zona geográfica. Fuente: Kumbure, Lohrmann, Luukka, and Porras (2022).	12
3.1.	Diagrama del trabajo	15
3.2.	Rolling method	17
4.1.	Diagrama de la metodología para LPC	25
4.2.	Diagrama de la metodología para la Predictibilidad	27
4.3.	Diagrama de la metodología para la simulación de trading	27
4.4.	Diagrama de la metodología para determinantes de predictibilidad	29
5.1.	Ejemplo de resultados obtenidos en acción BKNG.	32
5.2.	Combinaciones posibles que compusieron la grilla en su versión inicial	33
5.3.	Métricas obtenidas de la grilla para cada combinación, activo BKNG	34
5.4.	MAE optimizado para cada acción en al grilla	35
5.5.	Discriminante 1 optimizado para cada acción en al grilla	36
5.6.	Directional Accuracy optimizado para cada acción en al grilla	37
5.7.	Acciones optimizadas para cada métrica para el set 1	38
5.8.	Acciones optimizadas para cada métrica para el set 2	39
5.9.	Representación de señales por su dirección	40
5.10.	Métricas obtenidas gracias a las representaciones, comparadas con el DA óptimo.	41
5.11.	DA vs DIS	42
5.12.	DA vs MAE	43
5.13.	MAE vs DIS	43
5.14.	Relación entre Métricas Set 1.	45
5.15.	Relación entre Métricas Set 2.	45
5.16.	Cantidad de quiebres contra las distintas métricas.	51
5.17.	Ejemplo de histogramas para BKNG	52
5.18.	Promedio de rachas consecutivas para el S&P500	53
5.19.	Rachas de largo 1 vs Métricas.	54
5.20.	Promedio de rachas consecutivas para el S&P500 en el set 1	55
5.21.	Promedio de rachas consecutivas para el S&P500 en el set 2	55
5.22.	Logaritmo natural del volumen total vs DA, para cada acción	56
5.23.	Logaritmo natural de la variación del volumen vs DA, para cada acción	57

5.24.	Logaritmo natural de la variación del volumen vs DA, para cada acción en el set 1	57
5.25.	Logaritmo natural de la variación del volumen vs DA, para cada acción en el set 2	58
C.1.	Cantidad de quiebres contra las distintas métricas.	71
C.2.	Cantidad de quiebres contra las distintas métricas.	72

Capítulo 1

Introducción

Dentro del campo financiero, un constante desafío muy renombrado entre diferentes actores como inversores y profesionales del área corresponde al análisis de las fluctuaciones en la Bolsa de Valores, donde se ha abordado de diversas maneras para conocer el comportamiento y cómo este puede cambiar en el futuro. Es decir, ¿de qué manera se pueden determinar los futuros movimientos de las acciones?

La predicción del precio de las acciones es la tarea de estimar el valor del precio futuro de las acciones de una empresa [1]. Con lo cual, se define como una tarea desafiante, puesto que los datos suelen ser ruidosos y no estacionarios [2]. Por ello y otros factores que intervienen en el mercado, es difícil predecir los precios.

En este sentido, varios métodos o enfoques han sido propuestos y aplicados basados en diferentes técnicas de predicción y datos a predecir [3]. Por el lado de los métodos, se incluyen análisis estadísticos, aprendizaje automático tradicional (machine learning) y aprendizaje profundo (deep learning), entre otros; mientras que existe una variedad de datos que pueden ser trabajados y analizados según la técnica aplicada.

Si bien ya se han establecido diferentes modelos que son efectivos en la predictibilidad, se propone abordar el reto desde una nueva perspectiva con la finalidad de no solo predecir el precio de cierre de las acciones pertenecientes al S&P500, sino que también realizar un análisis de esta predictibilidad. El S&P500, llamado Standard and Poor's 500 s, es un índice bursátil ponderado de capitalización que incluye a las 500 principales empresas de Estados Unidos y mide su rendimiento correspondiente; este índice abarca múltiples sectores y se refiere como una “referencia de la salud general del mercado de valores” en el país, además que desempeña un papel vital en el mundo financiero.

Se optó por el mercado estadounidense por varias razones. Como se ha mencionado anteriormente, el índice S&P500 es uno de los mercados más grandes y líquidos a nivel mundial, además de contar con acciones de empresas líderes en diferentes sectores de la economía, brindando así mayor diversificación en oportunidades de inversión. Además de la estabilidad y transparencia, puesto que este país cuenta con una regulación financiera sólida y una vasta historia, entregando datos históricos fiables. En relación a estas razones, lo principal corresponde al reconocimiento de Estados Unidos como un referente global, en donde todo movimiento tiene un impacto significativo a nivel internacional y esto otorga una perspectiva

más amplia para comprender el mercado en términos generales.

En esta investigación, dicha problemática abordada se desarrolla en dos puntos claves. Primero, emplear un enfoque predictivo - Mediante LPC, método comúnmente utilizado para estimar señales de voz -, el cual pueda pronosticar el precio de cierre de diversos activos bursátiles pertenecientes al S&P500, con el fin de evaluar la predictibilidad de ciertos activos financieros. Segundo, estudiar los determinantes de la predictibilidad alcanzada para entender las principales características que permitan un mayor o menor grado de predictibilidad, puesto que al utilizar esta técnica en el área financiera, donde no había sido puesta a prueba anteriormente, podrá generar nueva información por evaluar y concluir, tal que permita clarificar el comportamiento de la predictibilidad en distintas acciones.

LPC (Lineal Predictive Coding), es una técnica que se ha utilizado tradicionalmente en el procesamiento de señales de voz, siendo propuesta su aplicación en un contexto diferente, como lo es dentro del campo económico y para la predicción de precios de acciones. Es relevante mencionar que la idea del usar LPC surge a partir de la observación de la similitud entre las señales de voz y los patrones de precios en el mercado de acciones. Este enfoque ha inspirado la idea de aplicar técnicas de análisis de señales como en este caso los es LPC.

Dentro del abanico de técnicas aplicadas en el ámbito financiero, ARIMA se destaca como una de las más ampliamente reconocidas y empleadas. ARIMA (Autoregressive Integrated Moving Average) se basa en la identificación de patrones temporales, estacionales y de tendencia en los datos históricos, lo que lo convierte en una herramienta sólida para el análisis de series temporales financieras. Por otro lado, LPC introduce una perspectiva innovadora al enfoque predictivo. Si bien ARIMA aborda la estructura temporal de los datos, LPC se enfoca en la extracción de características relevantes que no solo permiten la predicción, sino que también abren la puerta a una comprensión más profunda de los factores fundamentales que influyen en la predictibilidad del mercado financiero. Esta diferencia en enfoques proporciona un marco integral para explorar y comprender los movimientos del mercado desde distintas perspectivas analíticas.

El camino trazado de este estudio se articula en una serie de etapas interconectadas; lo primero es la recopilación de datos para su posterior limpieza, seguido de la implementación de LPC para analizar los aspectos de predictibilidad, simulación de trading y determinantes. Cada punto tiene una finalidad; en predictibilidad se analizan dos tipos (nivel y dirección), simulación de trading busca evaluar la rentabilidad de las acciones y si están afectadas por las métricas mencionadas previamente, mientras que se realiza una regresión lineal multivariable que busca encontrar los determinantes que mejor expliquen el comportamiento de la predictibilidad. Posteriormente, se establece comparaciones con ARIMA en términos de predictibilidad para tener un método de referencia validado.

De tal manera, la principal finalidad radica en obtener un análisis innovador sobre este tema; aunque ha sido tratado varias veces por numerosos artículos, esto sigue siendo un tema de alto interés que no ha logrado alcanzar una resolución definitiva hasta la actualidad. Además, se busca encontrar razones por las cuales ciertas acciones son más (o menos) predecibles que otras, considerando también la efectividad del enfoque propuesto. Asimismo, los datos provienen de la recopilación de los datos históricos de precios de acciones, específicamente

del índice S&P500 desde el 8 de Septiembre del 2010 al 8 de Agosto del 2022 y la aplicación del método LPC para generar predicciones de los precios futuros.

Ante esto, se presentan los objetivos de la investigación que guiarán el desarrollo de la misma:

- General:
 1. Evaluar la utilidad y efectividad de emplear LPC en el modelamiento de series de precios de acciones.
 2. Realizar una comparación exhaustiva entre el rendimiento del modelado mediante LPC y las técnicas tradicionales (ARIMA).
 3. Investigar la relación entre la predictibilidad y la rentabilidad, tanto en los métodos de modelamiento utilizados como en las métricas asociadas.
- Específicos:
 1. Comprender el funcionamiento teórico detrás de la aplicación de las herramientas.
 2. Implementar desde cero mediante Python las herramientas principales a utilizar.
 3. Obtener predicciones de diversas acciones bursátiles con las técnicas implementadas.
 4. Analizar los determinantes de la predictibilidad.
 5. Realizar simulación de trading con las predicciones obtenidas.

De igual forma, se busca aportar nuevas perspectivas y conocimientos al campo de la predicción de acciones. El uso del método LPC que no ha sido explorado en el contexto financiero, podría revelar nuevas formas de percibir la precisión de las predicciones y comprender de mejor forma el comportamiento del mercado bursátil. Además, se podrían tener implicaciones prácticas significativas en los resultados obtenidos para los diferentes actores del mercado e incluso proporcionarles una nueva herramienta adicional para la toma de decisiones.

La investigación se estructura en varias secciones. De manera inicial, el capítulo de antecedentes en la predicción de activos bursátiles, cumpliendo el rol de revisión literaria, consistiendo en la recopilación y análisis de estudios anteriores y teorías relevantes; seguido por el capítulo predicción lineal aplicada a activos financieros, con el rol de marco teórico, que proporciona la presentación de teorías, modelos y conceptos fundamentales. La metodología describe los procedimientos y técnicas utilizadas, posterior están los experimentos y resultados que detallan hallazgos obtenidos así como su interpretación, finalizando en la síntesis de las principales conclusiones y recomendaciones respecto al estudio realizado. Además, se agregan anexos con información adicional relevante para complementar la información y respaldar los análisis presentados.

Capítulo 2

Antecedentes en la Predicción de Activos Bursátiles

La precisión en la predictibilidad del comportamiento del mercado de valores se conoce como un desafío de suma importancia en el ámbito financiero. Diferentes actores de múltiples áreas buscan constantemente técnicas, herramientas y enfoques que permitan anticipar las fluctuaciones de precios de las acciones.

También hay otras variables que influyen en la predicción, sean factores macroeconómicos, desarrollo industrial, comportamiento de los inversores, eventos geopolíticos, tendencias y patrones técnicos. Esto entrega un panorama complejo y variable que disminuye el poder de predictibilidad efectiva en sí.

Ante esto, surge el objetivo de lograr un retorno ajustado por riesgo superior al mercado por parte de inversionistas institucionales como retailers que buscan anticipar el comportamiento del mercado [4]. Al contrario, este objetivo se presenta como lo opuesto a lo planteado por la hipótesis de mercado eficiente establecida por Fama en 1970 [5], la cual establece que en mercados eficientes, los precios de las acciones fluctúan a través de random walk (caminata aleatoria), y como tal los precios no pueden ser antedichos usando información existente, habría nula posibilidad de anticipar la dirección de los precios de las acciones como la magnitud de cambios [6].

En su propuesta, Fama propuso tres formas de eficiencia de mercado: Débil, postula que la información de los precios pasados ya se encuentra reflejada en el precio actual, siendo no factible predecir los precios futuros de las acciones y esta forma no supera la estrategia buy-hold en relación al retorno ajustado por riesgo. Por otro lado el Semi-Fuerte, los precios reflejan toda la información pública disponible que engloba tanto los precios pasados como aquella relacionada a indicadores macroeconómicos; donde una gestión activa que usa esta información no garantizaría superar el buy-hold consistentemente. Y por último el Fuerte, que indica que ninguna información puede ser usada para pronosticar el mercado.

De acuerdo a lo propuesto por Fama, algunos resultados empíricos sugieren la existencia de una eficiencia fuerte o semi-fuerte, ya que los resultados que se predijeron no se perciben favorables ante las expectativas de los inversionistas. En 2016, por ejemplo, Gerlein, Mc-

Ginnity, Belatreche, y Coleman se concluyó que los modelos de machine learning en lo que respecta su capacidad para predecir al menos la dirección del cambio de precio en el corto plazo variaba entre el 48 % y 54 % [7], donde se sitúa arriesgadamente similar a no ser más preciso que el simple lanzamiento de una moneda.

No obstante, se afirma haber predicho el mercado con un éxito considerable en la práctica de inversores y fondos de inversión al utilizar el análisis técnico [8]. Además que existe bastante evidencia que señala la existencia de capacidad predictiva en ciertos mercados y activos, sea diaria como intraday (intraday prediction).

En la tabla 2.1 se entrega un resumen de algunas investigaciones con las más altas capacidades predicativas utilizando diferentes técnicas de predicción en los últimos años.

Tabla 2.1: Publicaciones seleccionadas capacidad predictiva más elevada reportadas en la literatura.

<i>Research Title</i>	<i>Year</i>	<i>Author</i>	<i>Prediction Method</i>	<i>Markets</i>	<i>Best hit ratio</i>
<i>Forecasting stock indices: a comparison of classification and level estimation</i>	2000	<i>Mark T. Leung and Hazem Daouk and An-Sing Chen</i>	<i>Classification Model</i>	<i>US, UK, Japan</i>	<i>68 %</i>
<i>Forecasting stockmarket movement direction with support vector machines</i>	2005	<i>Wei Huang and Yoshiteru Nakamori and Shou-Yang Wang</i>	<i>Support Vector Machine</i>	<i>Japan</i>	<i>75 %</i>
<i>Predicting direction of stock price index movement using artificial neural networks and support vector machines</i>	2011	<i>Yakup Kara and Melek Acar Boyacioglu and Omer Kaan Baykan</i>	<i>Neural Network</i>	<i>Istanbul</i>	<i>75.74 %</i>
<i>Predicting the direction of stock market index movement using an optimized artificial neural network model</i>	2016	<i>Qiu Mingyue and Song Yu</i>	<i>Genetic Algorithm - Artificial Neural Network Hybrid Model</i>	<i>Japan</i>	<i>81.27 %</i>
<i>Predicting the direction of stock markets using optimized neural networks with google trends</i>	2018	<i>Jing Zhang and Shincheng Cui and Yan Xuand Qianmu Li and Tao Li</i>	<i>ISCA - BPNN Hybrid Model for Type II</i>	<i>USA - S&P500</i>	<i>86.81 %</i>
<i>Predicting the direction of stock markets using optimized neural networks with google trends</i>	2018	<i>Jing Zhang and Shincheng Cui and Yan Xu and Qianmu Li and Tao Li</i>	<i>ISCA - BPNN Hybrid Model for Type I</i>	<i>USA - DJIA</i>	<i>88.81 %</i>
<i>Forecasting and optimization stock predictions: varying asset profile, time window and hyperparameter factors</i>	2023	<i>Chaher Alzaman</i>	<i>LSTM</i>	<i>Canada - TSE</i>	<i>88 %</i>

En cuanto a las metodologías empleadas para desarrollar los modelos predictivos previamente mencionados, se construyen en base de un proceso que considera varios pasos. En general, lo necesario para comenzar es: data collection, data pre-processing (data cleaning) y definir el modelo de predictibilidad para trabajar.

En primer lugar, data collection se refiere al proceso de recopilación, sean datos e información relevante para algún uso en particular. Los datos financieros son un término general para una gama de datos involucrados en el sector de la industria financiera, incluidos datos de mercado, datos de empresas, índices de la industria, datos de precios, etc. [9], y lo cual en este contexto, representan el desempeño de las operaciones de la empresa, que pueden ser divididos en cuatro categorías:

- Datos de usuario: o conocidos como datos básicos, se refieren a la información personal del cliente (nombre, edad, género, entre otros). También incluyen clientes empresariales

(nombre, empresas afiliadas, monto de ventas y capital registrado). El desarrollo del producto y el apoyo a la toma de decisiones se llevan a cabo en función de otros datos de transacciones.

- Datos de transacciones de mercado: corresponde a toda actividad comercial en las plataformas de trading. Estos datos son más estandarizados y frecuentes, además que la cantidad de antecedentes es mayor y procesarla es más complejo, pero sus resultados poseen mayor información . Estos son utilizados para analizar los comportamientos de consumo para desarrollar productos y servicios; también para proporcionar servicios financieros en la cadena de suministro y establecer su ecosistema de comercializar productos.
- Datos de análisis: representan los resultados del análisis de la información de los datos originales. En estos corresponden al análisis de profundidad del modelo de negocios de la empresa, actividades, competencia y perspectivas. Sin embargo, el análisis de datos también implica un alto costo y toda metodología puede tener riesgo de sesgo o falta de generalización.
- Otros datos: en términos generales, incluye otros tipos de información que pueden considerarse necesarios en adición de los tres anteriores. También poseen un rol clave en el análisis de datos y toma de decisiones; además que incluyen datos relevantes de comportamiento, ubicación, negocios y procesos comerciales.

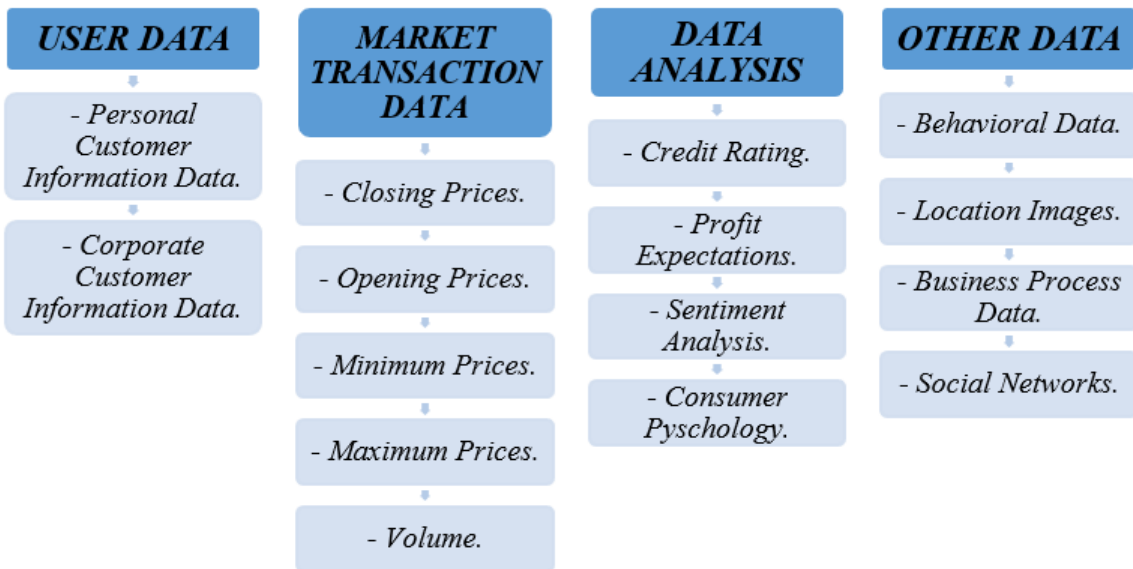


Figura 2.1: Clasificación de Información - Data Collection. Fuente: Zhang, Fan, Ye Ding, and Yuhao Liao (2022).

Por otro lado, dentro de la información más empleada corresponden a los datos de mercado en donde se incluyen precios de cierre, precios de apertura, precio mínimo, precio máximo y volumen. Más del 67 % de los estudios publicados en el año 2019 trabaja con este tipo de datos [10] - historical prices - y otros como texto e imagen, además de combinaciones entre

varios métodos.

En la Figura 2.2 se aprecia la distribución de combinaciones de tipos de variables de entrada utilizada, además de los casos donde solo se utiliza una:

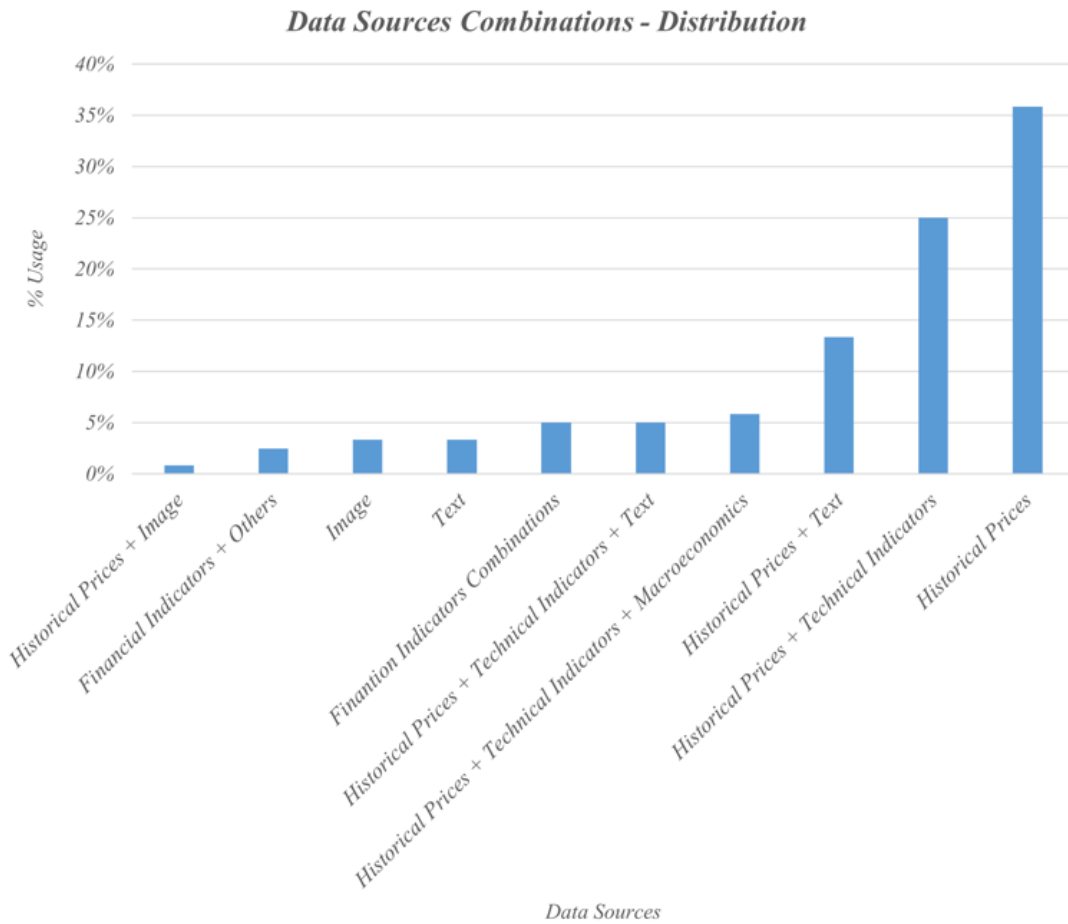


Figura 2.2: Distribución de las combinaciones de las fuentes de datos. Fuente: Jiang, W. (2021).

Adicionalmente, los datos financieros tienen ciertas características específicas que incluyen amplitud, exhaustividad, fiabilidad y continuidad. Otra característica corresponde a la granularidad de los datos o frecuencia en la variable a modelar; en la Figura 2.3 se muestra la granularidad de los datos utilizada en 138 artículos [6] de los cuales en su mayoría fueron publicados entre 2015 y 2019 y el resto fue publicado en las últimas dos décadas.

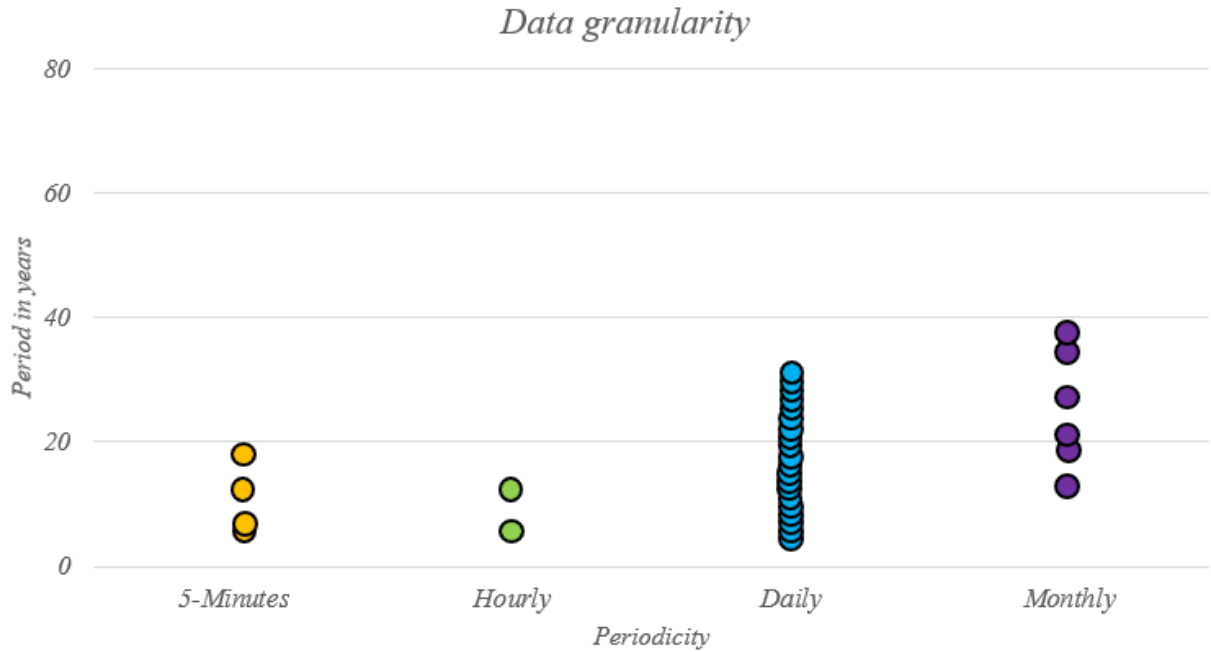


Figura 2.3: Granularidad de los datos utilizados en cada estudio. Fuente: Fuente: Kumbure, Lohrmann, Luukka, and Porras (2022).

El eje horizontal corresponde a la frecuencia de los datos, el número de círculos corresponde a la cantidad de publicaciones que hayan empleado dicha granularidad y el eje vertical la extensión de los datos de las publicaciones. La frecuencia más común de los estudios se centraron en las predicciones diarias con una extensión inferior a 30 años en general, seguido de mensuales por mayor extensión y de 5 minutos con una frecuencia menor.

Posterior del data collection, entra el pre-processing y en específico, el data cleaning, también conocido como limpieza de datos, es un paso crucial en el análisis de datos para generar resultados adecuados y este abarca el proceso de identificar y corregir/eliminar errores, inconsistencias y valores atípicos en un conjunto de datos. Dentro de sus técnicas, las más comunes son:

1. Normalización y estandarización de datos: verificación de uniformidad en la escala de los datos mediante la normalización de variables numéricas para estar dentro del mismo rango o estandarización de variables para tener una media cero y desviación estándar de uno.
2. Corrección de errores de formato: verificación y corrección de errores de formato en los datos. Sean tipográficos, problemas de codificación u otros.
3. Detección y manejo de datos duplicados: identificación y eliminación de registros duplicados en los datos.
4. Eliminación de valores atípicos: identificación y eliminación de valores atípicos que pueden sesgar los resultados del modelo.

5. Manejo de valores faltantes: manejar la falta de datos con técnicas como reemplazo de estos, medidas estadísticas de los otros datos existentes o eliminación de registros con valores faltantes según sea apropiado.
6. Filtrado y selección de variables relevantes: selección de variables por relevancia para el modelo y eliminación de aquellas sin aporte significativo.

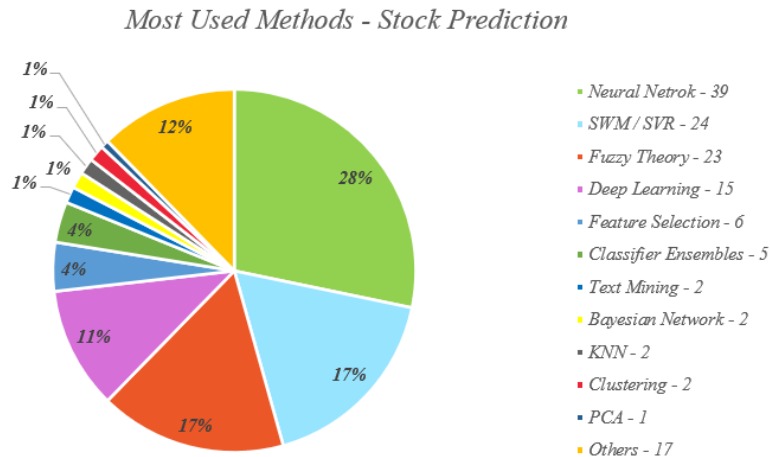


Figura 2.4: Diversidad de métodos utilizados.

Posteriormente, el punto determinante para trabajar la metodología restante es la definición del modelo de predictibilidad y sus pasos correspondientes. En la siguiente Figura 2.4 se muestran los métodos más empleados en la predicción de precios de acción. Esta investigación [6] abarcó 138 artículos que fueron publicados en los últimos 20 años, siendo en su mayoría publicados entre 2015 y 2019.

De tal manera, se aprecia que los modelos más empleados corresponden a Redes Neuronales (1 sola capa) con un 28 %, SVM/SVR y Fuzzy Theory con un 17 % en ambos y Deep Learning con un 11 %, mientras que en Otros se abarcan más métodos con un 12 %.

Otro punto se refiere a la evolución y cantidad de modelos de predicción desde el año 2000 hasta el 2019 [6]. Se menciona que lideraba el empleo de las ANN (Artificial Neural Networks) por el 2000 en donde sigue teniendo un camino sólido y fiable en su aplicación, seguido de las SVM (Support Vector Machine) que tomaron mayor uso en el 2009. Otros de menor uso son Clustering con su primer inicio en el 2008 y Bayesian Network con únicamente 4 apariciones entre 2012 y 2018, al igual que Decision Trees con la misma cantidad de apariciones entre 2010 y 2018.

Dentro de todos, el método Deep Learning tuvo una significativa presencia desde el año 2016 hasta el 2019 con 15 apariciones, seguido de Fuzzy con una presencia más antigua con 23 apariciones desde el 2002 y Feature Selection 14 apariciones desde el 2009. De tal manera, se evidencia que Deep Learning fue el principal método en los últimos años que se tomaron para la investigación de la evolución de los modelos (período 2000-2019); además que las

ANN y SVM se mantuvieron constantes en el tiempo siendo las más utilizadas en estos años.

Entre otros, se presentan los métodos más relevantes en machine learning:

- ARIMA (Autoregressive Integrated Moving Average): correspondiente a la combinación de tres componentes: autoregresión (AR), diferenciación (I) y promedio móvil (MA), siendo en conjunto un modelo estadístico para analizar y predecir series de tiempo. En general, es ampliamente utilizado en el análisis y predicción de series de tiempo en diversas áreas, como finanzas, demanda de productos, economía y otros gracias a su capacidad de comprender patrones temporales en los datos y realizar pronósticos precisos en base de estos patrones. Una reciente investigación [11] profundiza en el proceso de desarrollo de un modelo ARIMA para pronosticar los precios de las acciones, donde los hallazgos obtenidos demuestran que siguen siendo competitivos, demostrando que ha sido una técnica que se ha mantenido durante el tiempo.
- Redes Neuronales: estas funcionan con MLP (Multilayer Perceptron), algoritmo más renombrado en cuanto a la predicción de precios. Este algoritmo en sí es una red feed-forward con una capa de entrada, otra de salida y una amplia cantidad de capas intermedias, donde cada una incorpora capacidades de aprendizaje con datos varios. Adicionalmente esta red usa la técnica back-propagation en el set de training, y en este el error de predicción es propagado desde la capa de salida hacia la entrada modificando los pesos en cada capa.

En cuanto a este algoritmo, se ejemplifica mediante varios casos. En 2010, Dhar y socios emplearon MLP para predecir el precio de cierre del índice de la bolsa de India [12]. En paralelo, otros también lo usaron para fines similares. Si bien existieron buenos resultados en emplear MLPs, estas son demasiado sensibles a los parámetros de la arquitectura de la red.

Ante esto, han aparecido variaciones de modelos que emplean otras arquitecturas de redes neuronales, tales como combinación con algoritmos genéticos y otros para obtener mayor capacidad predictiva. Sin embargo, las redes neuronales presentan ciertas restricciones en los patrones de aprendizaje [12]. Además de que muestra soluciones locales y no globales, como espera; aparte que se ha conocido una limitación de overfitting [6].

- SVM (Support Vector Machine): en relación a lo anterior, surgen los SVM. Este algoritmo de aprendizaje supervisado se utiliza generalmente para resolver problemas de clasificación y regresión (support vector regression), en sí son métodos estadísticos de aprendizaje inteligentes.

El funcionamiento se basa en la idea de encontrar un hiperplano de máximo margen. Es decir, en un problema de regresión minimiza la distancia a los puntos observados en este y en uno de clasificación maximiza la distancia entre dos conjuntos de datos. Entre los algoritmos de clasificación lineal es la técnica más común gracias a los resultados que ha obtenido, siendo estos iguales o mejores que algoritmos más complejos [13].

- Deep Learning: o aprendizaje profundo, se enfoca en el entrenamiento de redes neuronales artificiales profundas para estudiar y comprender patrones complejos en conjunto de datos. Las redes neuronales profundas están compuestas por múltiples capas ocultas que procesan y transforman datos de manera repetida, extrayendo características de niveles de abstracción cada vez más altos.

Dentro de la predicción de precios y retornos de acciones, este método se describe como novedoso en donde hace menos de 10 años aparecen los primeros estudios de su aplicación en este contexto. Según lo planteado en la investigación de Kumbure y colaboradores en 2022 [6], el primer estudio fue en 2017 por Chong.

- Fuzzy Logic: o lógica difusa, es un enfoque de la lógica matemática que permite lidiar con la incertidumbre en la toma de decisiones. A diferencia de la clásica, esta permite la representación y manipulación de grados de verdad en lugar de valores binarios, teniendo un intervalo de resultados. En 2016, Ghanavati y su equipo propusieron un modelo combinando Fuzzy Clustering y Support Vector Machines, el cual superaba los modelos SVM [13].
- Algoritmos Genéticos: método inspirado en la teoría de la evolución y la genética, corresponden a una técnica de optimización y búsqueda. Para encontrar la mejor solución y maximizar la precisión de los modelos, imita los procesos de evolución natural. Hu and Li en el 2017 usaron este tipo de algoritmos para predecir el movimiento de 10 acciones de la bolsa de Taiwan y establecieron comparaciones con modelos benchmarks regresión lineal y regresión logística, de lo cual concluyeron que los algoritmos genéticos obtuvieron mayor precisión que los benchmarks [13].

En términos geográficos, el número de publicaciones de predicción de precios y retornos de acuerdo al país se muestra en la Figura 2.5. Asia lidera con 119 apariciones seguido de Norteamérica con 63 registros. No obstante, Estados Unidos presenta la mayor cantidad de publicaciones seguido de Taiwán y China con 28 y 25 publicaciones respectivamente. En cuanto al sector sudamericano, Brasil es el único presente en representación con 7 registros.

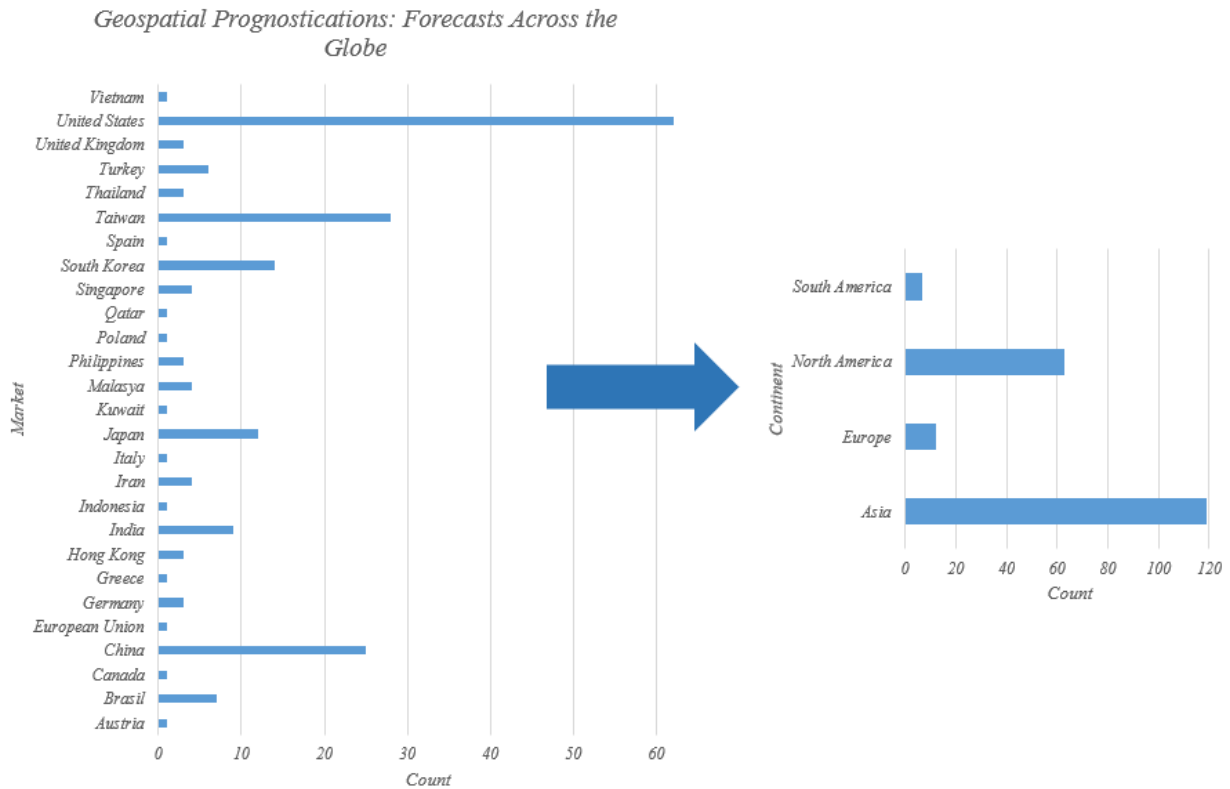


Figura 2.5: Publicaciones de predicciones de acuerdo a la zona geográfica.
Fuente: Kumbure, Lohrmann, Luukka, and Porras (2022).

Respecto a las variables modeladas, Estados Unidos es el país con mayor representación donde el índice con mayor interés corresponde a S&P 500 con 50 menciones, seguido de Nasdaq y Dow Jones con 14 y 13 apariciones respectivamente. Asia como un fuerte líder en el área también, sus índices más modelados son Taiex, Sse y Nikkei con 24, 14 y 12 artículos respectivamente.

Con respecto a la investigación presente, se persigue desarrollar una nueva aproximación para abordar este desafío mediante el enfoque del método LPC (Linear Predictive Coding). Es relevante mencionar que el uso del LPC dentro del contexto financiero es una propuesta que no ha sido ampliamente explorada. En este caso, se propone la combinación del manejo de señales eléctricas con el mercado financiero, idea que surge a partir de la observación de similitud entre las señales de voz y los patrones de precios en el mercado de acciones. Este enfoque ha inspirado la idea de aplicar técnicas de análisis de señales como LPC.

Mediante su aplicación, se busca generar nuevos conocimientos y resultados que puedan contribuir en la predicción del comportamiento del mercado financiero. Para respaldar esta propuesta, se examinaron ciertos estudios en cuanto a LPC.

Por ejemplo, Xu, de Fréin y Davis [14] proponen un nuevo método de filtro LPC para seguimiento de los cambios de frecuencia de SNR (Signal-to-Noise Ratio) bajas en tiempo real. Dicho nuevo método consiste en deslizar una ventana de análisis de longitud N muestras sobre la señal y aplicando este a los datos en ventana. Este propone un nuevo método robusto de resolución temporal para extraer y rastrear la frecuencia dominante desde señales multi-

frecuencia. Se obtienen tres conclusiones: (i) es un método que se resuelve en el dominio del tiempo (time-resolved method) y puede rastrear las variaciones de frecuencia en tiempo real, (ii) tiene capacidad de análisis en señales compuestas por múltiples frecuencias y (iii) puede identificar las características espectrales dominantes en entornos ruidosos.

Y en segundo lugar, en cuanto a la combinación de métodos, otro artículo [15] establece comparaciones entre una red basada en la atención (MHANet - Multi-head Attention Network) y una red convolucional temporal (TCN) para la mejora del habla (speech enhancement) basada en AKF (Augmented Kalman Filter). El método propuesto, DeepLPC-MHANet, produce estimaciones de LPC más precisas y logra puntajes de inteligibilidad y calidad más altos en comparación de otros métodos. Además, investigaciones futuras se direccionarán en el filtrado de Kalman para la mejora del habla en presencia de habla ruidosa y reverberante.

Otro punto relevante por mencionar corresponde al análisis de determinantes mediante la regresión multivariable. Este se define como un proceso estadístico para entender cómo una variable depende de otra; un ejemplo que representa el punto mencionado es un estudio aplicado en múltiples empresas inglesas, donde buscó sus determinantes mediante distintas variables de crecimiento, ganancias, estructura, etc.[16] Mostrando qué variables fueron las que tuvieron mayor significancia para explicar su comportamiento.

En relación al tipo de variables más importantes en la influencia del precio de las acciones, se han destacado:

Tabla 2.2: Tipos de variables y su literatura correspondiente

Type	Literature
Market Fundamentals	“Stock market predictability: A theoretical and empirical synthesis” by Campbell and Thompson (2008) [17]
Firm Characteristics	“Predictable dynamics in the cross-section of stock returns” by Bali, Brown, and Tang (2014) [18]
Technical Indicators	“Simple Technical Trading Rules and the Stochastic Properties of Stock Returns” by William Brock, Josef Lakonishok and Blake LeBaron(1992) [19]
Sentiment Indicators	“Investor Sentiment in the Stock Market” by Malcolm Baker and Jeffrey Wurgler(2007) [20]
Volatility	“Generalized Autoregressive Conditional Heteroskedasticity” by Tim Bollerslev(1986) [21]
Liquidity	“Illiquidity and stock returns: cross-section and time-series effects” by Yakov Amihud(2002) [22]
Behavioral Factors	“A Survey of Behavioral Finance” by Nicholas Barberis and Richard Thaler(2002) [23]

A través de un análisis exhaustivo de la literatura existente, se plantea la expectativa de obtener una visión más completa de los enfoques y resultados en el campo. Además, este análisis identifica el conocimiento actual y sienta las bases para la investigación propuesta, brindando una sólida base teórica y empírica para el desarrollo de la investigación.

Seguido de esto, se muestra que la hipótesis de mercado eficiente propuesta por Fama plantea que es imposible lograr retornos esperados superiores al mercado. Sin embargo, se contraponen evidencias que sí hay modelos con capacidad predictiva elevada. Asimismo, las técnicas más empleadas se enfocan en redes neuronales y SVM desde el comienzo de la predicción de los mercados, seguido de deep learning con una reciente entrada en el campo y

siendo la más empleada en el año 2019.

No obstante, también se revela la escasez de publicaciones en Sudamérica y las pocas que existen son provenientes de Brasil. Por otro lado, se presume que no existe una metodología establecida que indique cómo crear un sistema autónomo de trading [24].

El desafío de este estudio es implementar el método LPC en el contexto del mercado financiero. LPC es un subcaso de ARIMA, debido a que ambos son modelos autorregresivos (AR); solo que ARIMA posee adicionalmente un componente de diferenciación (I) y uno de memoria (MA). De tal manera, se buscó implementar LPC desde una perspectiva innovadora en un área que no ha sido explorado con anterioridad hasta el día de hoy, representando un avance en el estado del arte.

De tal manera, se ha impulsado el planteamiento de esta investigación para abordar una perspectiva que se manifiesta mediante la falta de exploración y aplicación del predictor LPC. Con el objetivo de avanzar en el estado del arte, se ha iniciado desde cero para implementación del método LPC enfocado en el mercado financiero. Seguido de ello, se adentra en el análisis de los determinantes de su predictibilidad, un aspecto que rara vez se aborda hasta la fecha. Además, se agrega un punto que enriquece la investigación para responder la pregunta “¿más predictibilidad se traduce en más rentabilidad?” al agregar simulación de trading. Con estas etapas relacionadas entre sí y su abordaje, se aspira a la contribución dentro del campo financiero.

Se establece LPC como una versión más simplificada de ARIMA, donde se busca poner a prueba la eficacia de este método y obtener resultados competitivos en la predicción de precios de acciones. Además, este enfoque puede proporcionar información clave a los actores del mercado y abrir nuevas oportunidades en el campo de la predictibilidad financiera.

De tal manera, las preguntas que busca resolver esta investigación son las siguientes:

1. ¿Existe capacidad predictiva del método LPC dentro del mercado financiero?
2. ¿El análisis nos podría indicar nuevas conclusiones sobre las razones por las cuales ciertos activos son más o menos predecibles?
3. ¿La existencia de predictibilidad significa obtener ganancias económicas en el trading?
4. ¿El método LPC tiene potencial para competir con las técnicas más avanzadas del estado del arte?

Capítulo 3

Predicción Lineal Aplicada a Activos Financieros

El contexto teórico se basó bajo el siguiente diagrama, siendo el hilo conductor del marco teórico. Al seguir esta representación gráfica permite una comprensión más profunda de la materia en cuestión:

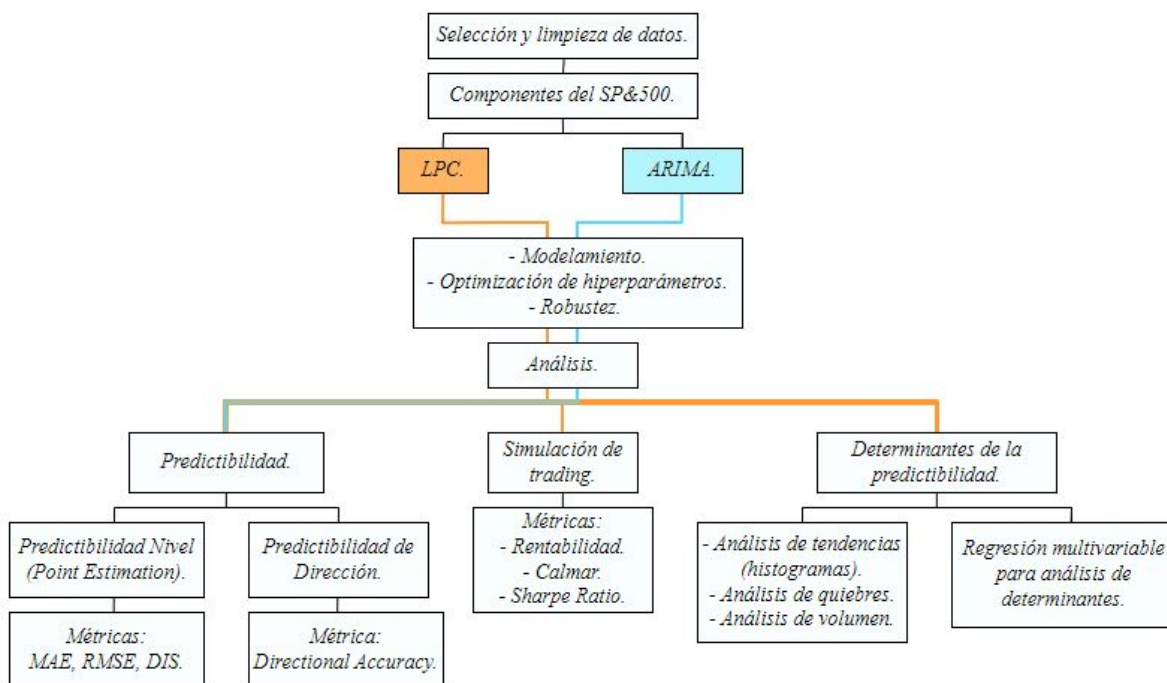


Figura 3.1: Diagrama del trabajo

3.1. LPC

Respecto al diagrama de la figura 3.1, para poder llevar a cabo las implementaciones de LPC se requiere la comprensión de los siguientes conceptos:

3.1.1. Teorema de Descomposición de Wold

El Teorema de Descomposición de Wold enuncia que una serie temporal de muestras correlacionadas puede ser generada a partir de una serie de muestras estadísticamente independientes (ruido blanco) y una serie determinística (la cual puede predecirse utilizando su información pasada) procesadas a través de un filtro lineal.[25]

3.1.2. Modelo Autorregresivo

Un modelo autorregresivo busca predecir la variable de interés utilizando una combinación lineal de los valores pasados de la misma variable. De este modo, un modelo autorregresivo de orden p , puede ser descrito como:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

con c una constante, ϕ los parámetros del modelo, y_t los valores en el tiempo t de la señal y ϵ_t ruido blanco. [26]

3.1.3. Análisis LPC

Linear Predictive Coding, o LPC por sus siglas en inglés. Para comprender cómo funciona, es necesario saber que existen métodos distintos de modelación que están basados en procesos estocásticos o procesos AR (autorregresivos).

La idea principal se basa en que la señal puede modelarse a través de una combinación lineal de p muestras anteriores más una señal de excitación o ruido blanco (error de predicción).

$$x[n] = \sum_{k=1}^p a_k x[n-k] + e[n] \quad (3.1)$$

Siendo similar a una regresión múltiple, pero con valores rezagados de y_t como predictores. De modo que la literatura se refiere a esto como un $AR(p)$ - modelo autorregresivo de orden p .

Donde:

- $x[n]$: Muestras de la señal.
- a_i : Coeficientes LPC.
- $e[n]$: Ruido blanco o error de predicción.

Para determinar los coeficientes LPC, es necesario utilizar las ecuaciones de Yule-Walker (basadas en el error cuadrático medio):

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} [x(n) - \sum_{k=1}^p a_k x(n-k)]^2$$

Luego se aplica $\frac{\delta E}{\delta a_k} = 0 | k = 1, \dots, p$

Donde quedan p ecuaciones lineales:

$$\sum_{n=-\infty}^{\infty} x(n-i)x(n) = \sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} x(n-i)x(n-k)$$

Luego se define $R(i) = \sum_{n=i}^{N-1} x(n)x(n-i)$
 Con lo cual la ecuación queda como:

$$\sum_{k=1}^p a_k R(i-k) = R(i)$$

que escrito en forma matricial:

$$RA = r$$

Teniendo R y r se obtienen los coeficientes LPC:

$$A = R^{-1}r$$

Finalmente el error se obtiene de 3.1 [26]:

$$e[n] = x[n] - \sum_{k=1}^p a_k x[n-k]$$

3.1.4. Método de Ventanas Deslizantes

También conocido como Rolling Analysis, sirve para observar la estabilidad del modelo en el tiempo. Una suposición común de modelos en series de tiempo se refiere a que los coeficientes son constantes con respecto al tiempo (estacionarios)[27].

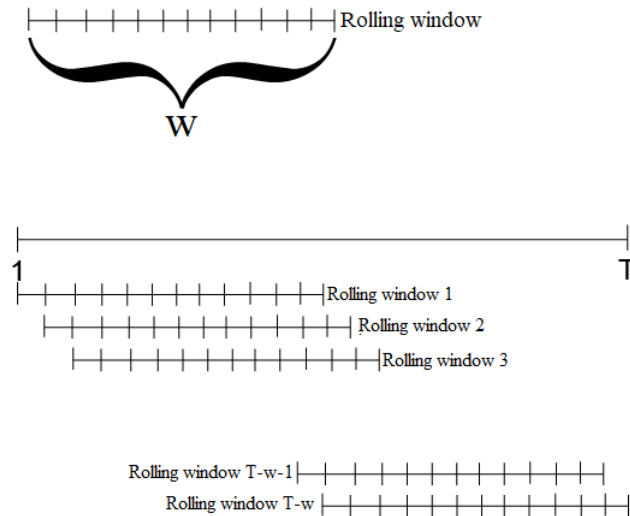


Figura 3.2: Rolling method

Para esto se elige un tamaño de ventana móvil, w , es decir, el número de observaciones consecutivas por ventana móvil. El tamaño de la ventana móvil dependerá del tamaño de la muestra, T y la periodicidad de los datos.

En general, puede usar un tamaño de ventana móvil corto para los datos recopilados en intervalos cortos y un tamaño mayor para los datos recopilados en intervalos más largos.

Si el número de incrementos entre las ventanas móviles es de 1, las ventanas totales corresponderían a $N_{tot} = T - w + 1$. Siendo la primera ventana desde la muestra 1 a la w , la segunda de la muestra 2 a $w+1$, y de esta forma sucesivamente.

3.2. Predictibilidad

Respecto al área de predicción las métricas utilizadas fueron:

3.2.1. Métricas

- Directional Accuracy:

Viene descrita por:

$$\frac{1}{N} \sum_t \mathbf{1}_{sgn(A_t - A_{t-1}) = sgn(F_t - A_{t-1})} \quad (3.2)$$

Donde A_t es el valor actual en el tiempo t , F_t es la predicción en el instante t , $sgn()$ es la función que extrae el signo del número y por ultimo $\mathbf{1}$ representando la función indicador, que toma el valor de 1 cuando ambos signos coinciden y 0 en cualquier otro caso. [28]

- MAE:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (3.3)$$

Mean Absolute Error, corresponde al promedio absoluto del error, siendo $|y_i - x_i|$ el error, donde y_i es la predicción y x_i el valor real.[29]

- Discriminante 1:

$$D1 = \frac{R_e(0)}{R_x(0)} \quad (3.4)$$

Con $R_e(0)$ La autocorrelación del error y $R_x(0)$ la autocorrelación de la señal. Representando la potencia del error dividido entre la potencia de la señal original.

- MAPE:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (3.5)$$

El error porcentual absoluto medio se compone por: A_t representa el valor real, F_t es el valor de la predicción y n es el número de observaciones.[30]

- MSE:

Error cuadrático medio viene dado por:

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} \quad (3.6)$$

Teniendo en cuenta que y_i es el valor real, \hat{y}_i es el valor de la predicción y n es el número de observaciones.[31]

3.3. Simulación de Trading

Se define como una herramienta que permite a los inversores y traders practicar estrategias de inversión y trading en un entorno virtual, replicando las condiciones del mercado real sin involucrar dinero real. Sus principales propósitos son: aprendizaje y práctica, evaluación de estrategias, análisis y pruebas de algoritmos, entre otros.

3.3.1. Métricas

La rentabilidad será analizada mediante:

- Calmar Ratio: el ratio de Calmar es un parámetro del rendimiento de inversiones que se define como el cociente entre la rentabilidad anualizada y el máximo drawdown del activo.

$$C_R = \frac{\text{Rentabilidad anualizada}}{|\text{Max(drawdown)}|}$$

Donde el drawdown se define como la diferencia entre el máximo de referencia de una curva y el mínimo, en un período de tiempo. Siendo una medida de riesgo que se utiliza en el ratio de Calmar, en lugar de la volatilidad.[32]

- Sharpe Ratio: es una medida para analizar el rendimiento de una inversión, comparado con el activo sin riesgo y ajustado por el riesgo que supone esa inversión. Para este caso en específico se definió como [33]:

$$S_R = \frac{\text{Retornos anuales}}{\text{Volatilidad anual}}$$

3.4. Determinantes de Predictibilidad

Conjunto de los factores, variables o condiciones que influyen en la capacidad de prever o anticipar una variable dependiente.

Dentro del mercado bursátil, estos determinantes pueden ser diversos y abarcar desde factores económicos y financieros hasta eventos geopolíticos, condiciones macroeconómicas, datos fundamentales de la empresa y otros elementos que afectan la dirección futura de los precios de los activos financieros. El análisis de los determinantes de predictibilidad busca comprender qué aspectos contribuyen a la capacidad de prever con precisión las tendencias y los movimientos del mercado.

3.4.1. CUSUM

Cumulative Sum o CUSUM [34] es un algoritmo que detecta los cambios en el promedio de una serie temporal que es estacionaria entre dos puntos. Una mejora de este algoritmo es

el Probabilistic CUSUM.

Si se definen observaciones arbitrarias en una secuencia temporal, se tiene:

$$Z_t = \frac{X_t - \mu_X}{\sigma_X}$$

Si se asume que se cumplen las condiciones para el teorema del límite central, siendo estas la suma de n variables aleatorias independientes, con media y varianzas finitas, implica que su función de distribución se aproxima a una distribución normal, se obtiene lo siguiente:

$$S_T = \sum_{t=1}^T Z_t \sim N(0, T)$$
$$\implies \tilde{S}_T = 1/\sqrt{T} S_T \sim N(0, 1)$$

$$\Phi(\tilde{s}_T) \approx P(\tilde{S}_T \leq \tilde{s}_T)$$

con Φ la CDF (Cumulative Density Function) de una distribución normal estandarizada.

El valor resultante se puede interpretar como la probabilidad de que la suma acumulada teórica sea tan pequeña como la que se está observando.

Para cada nuevo punto de datos se obtiene una medida de cuán extrema es la observación respectiva. Una vez que se supera un cierto umbral de 'improbabilidad', se marca como un punto de cambio y se reinicia el algoritmo.

3.4.2. Variables Económicas

- Bond Spread: se refiere al diferencial de tasas de interés entre bonos. Este concepto es una forma común que los actores del mercado comparan los valores entre bonos, al igual que las relaciones precio-beneficio se utilizan para las acciones; en términos matemáticos, el diferencial es la resta del rendimiento de un bono de otro [35]

Es un indicador relevante en el análisis y comparación de bonos en el mercado, este evalúa el riesgo crediticio y la percepción de los inversores en relación a la solvencia.

- Market Index: es una medida estadística utilizada para evaluar el desempeño general de un conjunto de acciones, bonos u otras inversiones dentro de un mercado financiero determinado. Estos índices son una referencia financiera para la evaluación del rendimiento general del mercado; también se tienden a agrupar por una o múltiples industrias como es el caso de S&P500, DJIA (Dow Jones Industrial Average) o Nasdaq, entre otros. Cada índice de mercado tiene una metodología determinada para derivar su valor, tiende a ser un tipo de promedio ponderado de los valores; este puede ser por capitalización de mercado, ingresos o precio de las acciones [36].
- Market Capitalization: Se define como el valor total de las acciones de una empresa en el mercado, siendo un método simple para evaluar el tamaño de una empresa y el riesgo de invertir en sus acciones. [37]

Es una métrica importante que clasifica a las empresas en función de su tamaño en el mercado. Se calcula multiplicando el número de acciones en circulación de una empresa por el precio actual de la acción.

- Volumen: en activos financieros, se refiere a la cantidad total de activos que se compran y venden en un mercado financiero durante un período de tiempo específico. Representa la magnitud de las transacciones y la liquidez del mercado [38].

Además, es una métrica importante en los mercados financieros dado que puede indicar la actividad y la participación de los inversores. Un alto volumen de negociación puede indicar un mercado activo y líquido, lo que puede facilitar la compra y venta de activos. Por otro lado, un bajo volumen de negociación puede indicar falta de interés o cautela por parte de los inversores, lo que podría afectar la liquidez del mercado.

- Price to Book per Share: es una medida financiera utilizada para evaluar si una acción está sobrevalorada o infravalorada según su valor contable, la cual es una herramienta relevante para analizar la salud financiera de una empresa desde una perspectiva contable. En cálculo, es la división del precio de mercado de una acción entre el valor libro por acción. Es decir, se refiere a la fracción del patrimonio de una empresa que se atribuye a cada acción ordinaria si la empresa se liquida. Dicha medida indica lo que obtendrían los accionistas si, teóricamente, vendieran todos sus activos y pagaran todos los pasivos de una empresa [39].

- Hurst Exponent Random Walk: es un índice de dependencia utilizado para analizar series temporales y determinar el comportamiento a largo plazo, ampliamente utilizado en la predicción y modelado de fenómenos financieros y económicos.

Dentro de los rangos, existen diferentes puntos e interpretaciones. Si el valor se encuentra de 0.5 a 1, indica una autocorrelación positiva a largo plazo, sugiriendo que un valor alto probablemente será seguido de otro valor alto sucesivamente; si el valor está entre 0 a 0.5, indica una autocorrelación negativa y no existe tendencia, donde un valor alto probablemente será seguido de uno bajo y vice versa. Si el valor es 0.5 o cercano, significa que la serie es completamente no correlacionada, tendiendo valores completamente aleatorios, denominándose random walk.[40]

- Trend Days of 30 Days ADX: se refiere al número de días en un período de 30 días donde el Average Directional Index (ADX) muestra una tendencia significativa en el mercado. El ADX se utiliza para cuantificar la fuerza de las tendencias; sus cálculos se basan en una media móvil de la expansión del rango de precios durante un período de tiempo determinado - 30 días en este caso [41]. Sus valores ayudan a identificar las tendencias más sólidas para operar, además de su relevancia para distinguir entre condiciones con y sin tendencia.
- VIX Index: el VIX tiene como objetivo evaluar la magnitud de los cambios de precios del S&P 500, o su volatilidad. Cuanto más intensas sean las fluctuaciones de precios en el índice, mayor será la cantidad de inestabilidad, y viceversa. Además de ser un índice para medir la volatilidad, los comerciantes también pueden negociar futuros, opciones o especular sobre los cambios en la volatilidad.

La volatilidad se midió de 2 formas distintas: El primer método utiliza cálculos estadísticos sobre precios anteriores durante un período de tiempo determinado para analizar

la volatilidad histórica. El proceso implica el cálculo de varios números estadísticos en los conjuntos de datos de precios históricos, incluida la media, la varianza y, finalmente, la desviación estándar.

El segundo, implica inferir su valor de los precios de las opciones. Las opciones son instrumentos derivados cuyo precio depende de la probabilidad de que el precio actual de una acción en particular se mueva lo suficiente como para alcanzar un nivel particular[42].

- **Analysts:** un analista bursátil es un tipo de analista financiero que hace predicciones sobre el desempeño del mercado de valores, generalmente centrándose en las acciones en las que invierte una empresa. Como analista de acciones, debe investigar empresas o industrias específicas, realizar análisis financieros basados en datos históricos y tendencias del mercado, y crear modelos para predecir el desempeño futuro. Además, basa el rendimiento anticipado de una acción en los informes técnicos que incluyen estimaciones financieras, proyecciones y recomendaciones [43].
- **Put Call Ratio:** indicador comúnmente utilizado para determinar el estado de ánimo (sentimiento general) de un mercado. Un "put" se refiere a una opción de venta, siendo un derecho a vender un activo a un precio determinado. Un call es una opción de compra, significando el derecho a comprar un activo a un precio determinado [44]. De tal manera, la relación put-call se calcula dividiendo el número de opciones de venta negociadas por el número de opciones de compra negociadas.
- **Bid Ask Spread:** Bid consiste en el precio más alto que un comprador está dispuesto a pagar por un número de acciones y Ask se refiere al precio más bajo que un vendedor está dispuesto a vender su acción. Por lo que el diferencial de oferta y demanda es esencialmente la diferencia entre el precio más alto que un comprador está dispuesto a pagar por un activo y el precio más bajo que un vendedor está dispuesto a aceptar..[45]

3.4.3. Regresión Lineal Múltiple

Posteriormente, para obtener un análisis de determinantes e identificar cuáles variables de las mencionadas anteriormente poseen una mayor importancia a la hora de explicar la predictibilidad se aplican regresiones lineales.

También conocida como RLM, es un enfoque estadístico que permite generar un modelo en el cual el valor de una variable dependiente (Y) se determina a partir de un conjunto de variables independientes llamadas predictores (X1, X2, X3, etc.) [46]. Estos pueden emplearse para predecir el valor de la variable dependiente o evaluar la influencia que tienen los predictores sobre esta.

3.4.3.1. Condiciones

Se deben cumplir ciertas condiciones para obtener resultados válidos y confiables [46].

- **No Colinealidad:** colinealidad se refiere a la situación donde dos o más predictores están altamente correlacionados entre sí. En cuanto a la No Colinealidad, es la situación donde las variables independientes (predictores) no están correlacionadas y como tal, no existe una fuerte relación lineal entre estas. Esto es el escenario deseable para obtener resultados confiables y evitar multicolinealidad.

- Parsimonia: el principio busca la simplicidad en la construcción de modelos, lo que implica utilizar la menor cantidad de predictores para explicar la variabilidad observada en la variable dependiente. Es decir, implica seleccionar las variables relevantes y eliminar aquellas que no influyen realmente en la predicción del fenómeno.
- Relación lineal entre los predictores numéricos y la variable respuesta: cada predictor numérico debe estar linealmente relacionado con la variable dependiente Y , mientras los demás predictores se mantienen constantes. La verificación recomendada es examinar los residuos del modelo según cada predictor, puesto que cuando la relación es lineal los residuos se distribuyen aleatoriamente cerca de cero.
- Distribución normal de los residuos: los residuos deben estar distribuidos de forma normal con media cero. Se tiende a comprobar mediante histogramas, cuantiles normales y otros tests.
- Variabilidad constante de los residuos: la situación se denomina homocedasticidad, siendo una propiedad deseable en el análisis de regresión para su interpretación y fiabilidad. Se refiere a la constancia de la varianza de los residuos en los valores de las variables independientes. Cuando se cumple esta propiedad, los residuos están distribuidos de manera constante alrededor de cero en toda la distribución de los valores predictores.
- No autocorrelación: las observaciones del modelo deben ser independientes entre sí. Es decir, no debe haber relación sistemática entre los residuos para evitar sesgos.
- Valores atípicos, con alto leverage o influyentes: también referidos como outliers, son observaciones que desvían significativamente el patrón de los datos en conjunto. Es relevante la identificación de observaciones que sean atípicas que puedan influir en el modelo, donde la forma más sencilla es mediante los residuos.
- Tamaño de la muestra: en sí no se trata de una condición, pero es crucial para la obtención de resultados fiables. Si la muestra es insuficiente, posiblemente los predictores no influyan en tener una significancia en el modelo.

Según el libro Handbook of Biological Statistics, sugiere que el número de observaciones sea mínimo entre 10 y 20 veces el número de predictores del modelo.

3.4.3.2. Pruebas Estadísticas

Para verificar ciertas suposiciones o propiedades de los datos y modelos, se aplican ciertas pruebas que tienen un propósito específico. Además, su aplicación adecuada es clave para obtener resultados confiables.

A continuación, se detallan los aplicados:

- Test Durbin-Watson [47]: evalúa si existe autocorrelación (correlación entre los residuos en diferentes períodos de tiempo) en los residuos del modelo de regresión.

La hipótesis nula (H_0) indica que no hay autocorrelación en los residuos, la cual debe ser aceptada y se evalúa de acuerdo al estadístico DW que toma valores entre 0 y 4, donde si es cercano a 2 sugiere que H_0 se confirma. Y en nivel de significancia, el valor p debe ser superior a 0,05 (5%). De caso contrario, se rechaza la hipótesis nula.

- Test Shapiro-Wilk [48]: se aplica para verificar si los residuos siguen una distribución normal. Presenta dos hipótesis, la nula plantea que los datos tienen distribución normal y la alternativa indica lo contrario, donde no hay.

En este caso, el nivel de significancia indica la probabilidad. Si el valor p es mayor que el nivel de significancia ($p > 0.05$) se confirma la hipótesis nula, siendo este el escenario deseable.

- Test Breusch-Pagan [49]: detecta la heterocedasticidad (los residuos tienen varianzas desiguales) en los residuos del modelo. Es decir, identificar si la varianzas de los residuos no es constante.

Su hipótesis nula plantea que no hay heterocedasticidad en los residuos, implicando que la varianzas de los residuos es constante. En este caso, si el valor p es mayor que el nivel de significancia se asume H_0 e indica homocedasticidad.

- AIC: es un método matemático el cual evalúa qué tan bien se ajusta un modelo a los datos a partir de los cuales se generó. En estadística, AIC se utiliza para comparar diferentes modelos posibles y determinar cuál se ajusta mejor a los datos. A menor valor de AIC mejor es el modelo, y este se calcula a partir de:

- El número de variables independientes utilizadas para construir el modelo.
- La estimación de máxima verosimilitud del modelo (qué tan bien el modelo reproduce los datos).

El modelo de mayor ajuste según AIC es el que explica la mayor cantidad de variación utilizando la menor cantidad posible de variables independientes[50].

Capítulo 4

Metodología

Posterior al contexto teórico, el procedimiento realizado fue el siguiente:

La implementación del trabajo se realizó de manera íntegra en Python y adicionalmente, se utilizó Google Colaboratory debido a lo que abarca el proyecto.

La metodología general del proyecto viene dada por la figura 3.1. Por lo que en esta sección se entra en detalle a cada sección del diagrama.

4.1. Metodología para la implementación de LPC

Al profundizar en los detalles de la implementación y programación, se empleó la siguiente metodología:

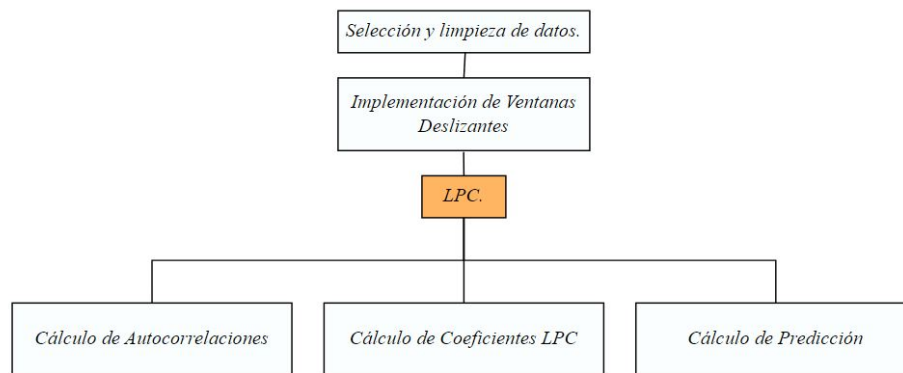


Figura 4.1: Diagrama de la metodología para LPC

Para la implementación de algunos pasos fue necesario realizar un procedimiento más extenso. Este se explica paso a paso a continuación:

4.1.1. Limpieza de datos

Como se indicó anteriormente, el dataset pertenece al S&P500 y va desde el 8 de Septiembre del 2010 al 8 de Agosto del 2022.

Los datos se sometieron a pre-procesamiento, dado que al momento de obtenerlos - 500 acciones - tenían datos relacionados a Date, High, Close, Low, Open y Volume, en donde los que finalmente se utilizaron las columnas Date, Close y Volume. Esto se consideró debido al motivo que se buscó predecir el precio de cierre diario y el volumen para un posterior análisis.

Debido a esto, se modificaron los dataset para que únicamente contengan las acciones con las columnas mencionadas anteriormente; luego se procedió a ver la cantidad de muestras que contenía cada acción incluida. Se examinaron en profundidad los datos, revelando que no todas las acciones contaban con un total de 3.000 muestras disponibles, puesto que algunas empresas entraron después del inicio del período de recolección de datos, mientras que otras se retiraron antes del final. Por ello, se decidió eliminar todos aquellos activos que no tuvieran las 3.000 acciones bursátiles a analizar con el fin que estuvieran en las mismas condiciones las que se utilizarían.

Sujeto a lo anterior, resultaron 437 acciones disponibles.

4.1.2. LPC

El proceso de implementación del predictor se dividió en etapas:

1. Proceso de enventanado, donde se aplicó el método rolling o de ventanas deslizantes, separando así la señal en $3000 - w$ ventanas, con w siendo el largo de la ventana o la cantidad de muestras que esta tenga.
2. A cada ventana se le aplica MVN (Mean-Variance Normalization). Guardando el promedio y desviación de cada ventana.
3. Se aplica LPC a cada ventana, eligiendo un valor p el cual representa la cantidad de rezagos o muestras a utilizar para realizar los cálculos del análisis LPC y obtener la predicción. Con esto se obtiene la predicción de la muestra $w+1$, es decir, el valor siguiente de la ventana.
4. Al realizar el paso anterior en todas las ventanas, se crean los vectores para calcular las métricas, los cuales eran; el vector que contiene todas las predicciones y el vector que contiene todos los valores reales, respectivos a cada predicción.
5. Se calculan todas las métricas mencionadas en el capítulo 2.
6. Para visualizar y tener una mayor comprensión de los resultados, se revierte la normalización y se obtienen los valores de los precios de cierre predichos.
7. Se repitió lo anterior para las 437 acciones bursátiles.

4.2. Metodología para la Predictibilidad

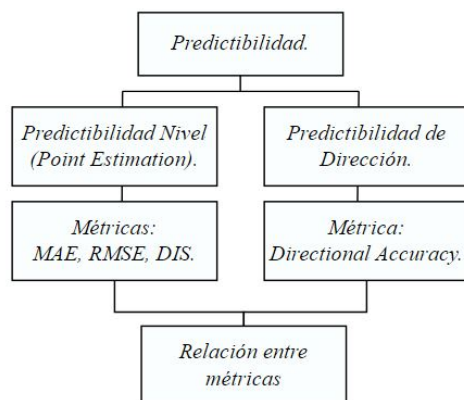


Figura 4.2: Diagrama de la metodología para la Predictibilidad

Constó de realizar las predicciones y una vez obtenidas estas, observar cómo se comportaban entre sí.

4.3. Metodología para la simulación de trading



Figura 4.3: Diagrama de la metodología para la simulación de trading

4.3.1. Estrategias

La simulación se define como una prueba de retrospectiva o también llamada backtesting, la cual se realizó con la librería Signal Backtester. Uno de los supuestos de la operación backtesting señala que en cada operación, sea compra o venta, se invierte el 100 % del patrimonio.

Para este procedimiento se utilizó una librería que tiene distintas estrategias para testear los modelos [51], los cuales son:

- **Two side stoploss takeprofit reversed:** abre posiciones tanto para compra como para venta, y cierra una posición con stoploss o take profit para una posición abierta de compra. Adicionalmente, si una posición de compra está abierta, y se activa una señal de venta se cierra la posición abriendo otra nueva.
- **Two side stoploss trailing reversed:** cumple con la tarea de abrir posiciones tanto para compra como para venta, y cierra posición con stoploss. El stoploss es dinámico cambia a medida que el precio cambia. Adicionalmente, si una posición de compra está abierta, y se activa una señal de venta se cierra la posición abriendo otra nueva.
- **One side buy stoploss takeprofit:** se refiere al acto de apertura de la posición sólo para compra. Esta posición se cierra sólo con la activación del stoploss o takeprofit
- **One side sell stoploss takeprofit:** abre la posición sólo para venta; esta se cierra con sólo la activación del stoploss o takeprofit.
- **One side buy stoploss trailing:** ejecuta la apertura de la posición sólo para compra. Esta posición se cierra con la activación stoploss. Esta estrategia cierra posición con stoploss. El stoploss es dinámico cambia a medida que el precio lo hace.
- **One side sell stoploss trailing:** esta estrategia abre la posición sólo para venta. Esta posición se cierra con el stoploss. Esta estrategia cierra posición con stoploss. El stoploss es dinámico cambia a medida que el precio también lo hace.

Finalmente los parámetros utilizados, se basaron en un estudio reciente que utilizó la misma librería.[33]

- Saldo inicial: 10,000,000
- Comisión: 0.05 %
- Takeprofit: 50 %. Se escogió este valor elevado para evitar que la operación se cierre por el concepto take profit.
- Stoploss: 50 % Se escogió este valor elevado para evitar que la operación se cierre por el concepto de stop loss.
- Trailing Stop: 3 %

4.4. Metodología para el Análisis de Determinantes

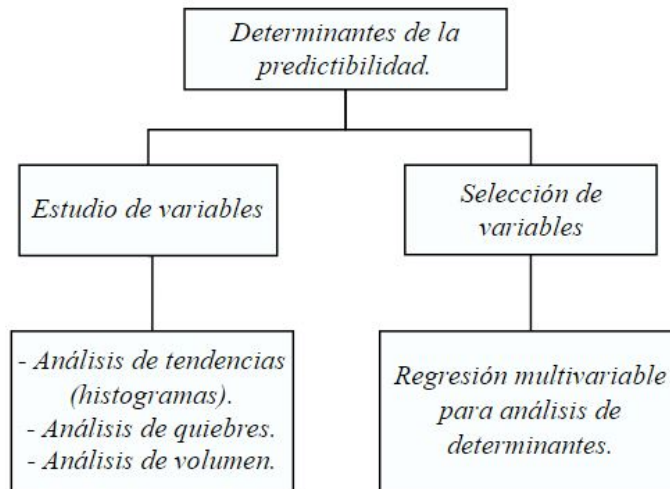


Figura 4.4: Diagrama de la metodología para determinantes de predictibilidad

4.4.1. Tipos de Variables

Los tipos de variables fueron escogidos por la literatura, donde se encontraron las siguientes:

- **Market Fundamentals:** se refieren a un conjunto de variables y factores que son considerados esenciales para comprender la salud y el desempeño general de un mercado financiero, una industria o una economía en particular. Estas variables suelen proporcionar una base sólida para tomar decisiones de inversión y evaluar la dirección en la que se mueve un mercado[52].
- **Firm Characteristics:** se utilizan para medir las características de una empresa en particular. Estas variables pueden incluir la internacionalidad, el apalancamiento financiero, el tamaño de la empresa, la rentabilidad, la intensidad de capital, la industria, el crecimiento, entre otros. Las Firm Characteristics son importantes para el análisis fundamental de una empresa, ya que proporcionan información sobre su salud financiera, su posición en el mercado y su capacidad para competir en su industria. Además, estas variables pueden ser útiles para identificar oportunidades de inversión y para desarrollar estrategias de marketing efectivas [16]
- **Technical Indicators:** son utilizadas en el análisis técnico de los mercados financieros. Estas variables se basan en el análisis de gráficos y patrones de precios para predecir futuros movimientos del mercado [53].
- **Sentiment Indicators:** miden y analizan el sentimiento o la opinión de los participantes del mercado. Estas variables se basan en el análisis de las emociones y las percepciones de los inversores, traders y otros actores del mercado.[54]

- Volatility: medida de la amplitud de los cambios en el precio de un activo financiero durante un período de tiempo determinado. La volatilidad es una variable que se utiliza para medir el riesgo de un activo financiero o un portafolio, ya que cuanto más volátil sea, mayor será el riesgo de pérdida, pero también mayores serán las expectativas de rendimiento.[55]
- Liquidity: evalúan la capacidad de una empresa o un mercado para cumplir con sus obligaciones financieras y para medir la facilidad de convertir activos en efectivo. La liquidez es esencial para mantener la estabilidad financiera y facilitar las transacciones comerciales[56]
- Behavioral Factors: se relacionan con el comportamiento y las decisiones de los inversores y participantes del mercado. Estos factores tienen en cuenta aspectos psicológicos y emocionales que pueden influir en las decisiones de inversión. Son importantes para comprender cómo las emociones y los sesgos cognitivos pueden influir en los mercados financieros y en las decisiones de inversión. Los inversores y analistas utilizan estos factores para evaluar el sentimiento del mercado y tomar decisiones informadas sobre la compra y venta de activos[57].

4.4.2. Selección de Variables

Las variables utilizadas para este experimento:

Tabla 4.1: Variables utilizadas para el análisis de determinantes

NAME	TYPE
Bond Spread Beta with Assets	Market Fundamentals
Market Index Beta with Assets	Market Fundamentals
Short Term Mean of Rolling STD	Risk Measure and Volatility
Long Term Mean of Rolling STD	Risk Measure and Volatility
Market Cap Rolling Mean	Firm Characteristics
Volume Coeficient of Variation	Firm Characteristics and Liquidity
Price to Book per Share Rolling Mean	Stock Behavior
Streak Histogram	Stock Behavior
Hurst Exponent Random Walk Days	Stock Behavior
Number of CUSUM Breaks of Price	Signal Behavior
Trend Days of 30 Days ADX	Technical Indicator
VIX Index Beta	Sentiment Indicator and Volatility
Analyst Dispersion	Sentiment Indicator and Volatility
Analyst Number	Sentiment Indicator
Put Call Ratio Rolling Mean	Sentiment Indicator
Bid Ask Spread Rolling Mean	Market Fundamentals and Liquidity

Las variables fueron obtenidas mediante Refinitv Eikon Datasets, donde se extrajo la serie temporal de las variables. En cuanto a las que fueron calculadas, son las que poseen beta, promedios y desviaciones en ventanas deslizantes.

- Bond Spread Beta with Assets: B se calcula utilizando Bond Spread y los retornos de las acciones, de la siguiente manera:

$$B = \frac{Cov(r_a, r_b)}{Var(r_b)}$$

Con r_a los retornos de la acción y r_b el diferencial de bonos.

- Market Index Beta with Assets: se refiere al mismo cálculo del B anterior, pero siendo este entre el índice de retornos del S&P500 y los retornos de la acción en particular.
- Short Term STD: desviación estándar de los precios, con ventanas deslizantes de 21 días.
- Long Term STD: desviación estándar de los precios, con ventanas deslizantes de 252 días.
- Market Capitalization Rolling Mean: se refiere a la capitalización de mercado en promedio obtenido mediante ventanas deslizantes.
- Volume Coefficient of Variation: es el obtenido de

$$CV = \sigma_V / \mu_V$$

con σ_V la desviación del volumen y μ_V el promedio del volumen.

- Streak Histograms: rachas de histogramas promedio del precio de cierre de la acción obtenido del experimento en la sección 5.5.2
- Hurst Exponent Random Walk: obtenido del precio de cierre del activo.
- Number of CUSUM Breaks: número de quiebres del activo, utilizando el método CUSUM explicado en la sección 3.4.1.
- Trend Days of 30 Days ADX: el ADX se utiliza para medir la fuerza de las tendencias. Los cálculos del ADX se basan en una media móvil de la expansión del rango de precios durante un período de tiempo determinado, Para este caso 30 días. [41]
- VIX Index Beta with asset: cálculo del Beta entre el índice VIX es el índice de volatilidad (Volatility Index), con el activo.
- Analyst Dispersion: coeficiente de variación, para este caso la desviación estándar de las ganancias por acción estimada por analistas obtenidas del promedio de ventanas deslizantes.
- Analyst Number Rolling Mean: cantidad de analistas calculados con el promedio de ventanas deslizantes.
- Put Call Ratio Rilling Mean: obtenido del promedio de ventanas deslizantes.
- Bid Ask Spread Rolling Mean: Bid Ask Spread obtenido del promedio de ventanas deslizantes.

Finalmente se utilizó el criterio de información de Akaike (AIC) para elegir el mejor modelo.

Capítulo 5

Experimentos y Resultados

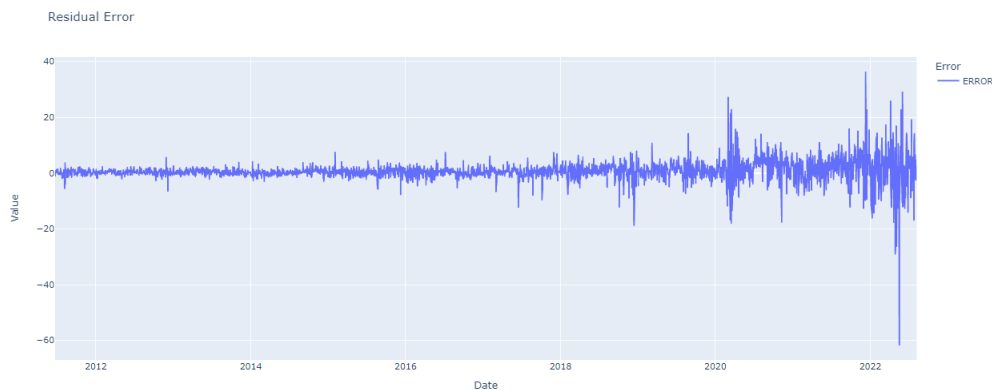
Siguiendo el orden del diagrama mostrado en la figura 3.1 se presentarán los resultados obtenidos:

5.1. LPC, Modelamiento, Optimización y Robustez

Los primeros resultados fueron las predicciones obtenidas mediante el predictor LPC, donde se logró visualizar la señal original, la predicción y el error residual.



(a) Comparación de señal original y predicción en una ventana.



(b) Error residual a lo largo de la serie de tiempo

Figura 5.1: Ejemplo de resultados obtenidos en acción BKNG.

Como se pudo observar en la figura 5.1, la predicción posee el mismo comportamiento y forma que la señal original. Sin embargo, está levemente desplazada hacia la derecha, como una especie de retraso.

En cuanto al error residual, se observó que este aumentaba a medida que avanzaba la serie de tiempo. Este comportamiento se debe a que a medida pasaba el tiempo, aumentaba el valor del precio en las acciones y sus fluctuaciones eran de mayor tamaño.

5.1.1. Optimización: Búsqueda de hiperparámetros

Con el fin de optimizar la predicción de LPC, se realizó un algoritmo de fuerza bruta. La finalidad consiste en encontrar los mejores hiperparámetros (w : tamaño de ventana y p : número de rezagos), el cual consistió en la creación de una grilla con distintas combinaciones para diferentes p y w .

Teniendo que p y w pudiesen tomar los siguientes valores:

$$p = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 60, 80, 100, 120]$$

$$w = [30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250, 300, 350, \dots, 400, 450, 500, 550, 600]$$

p	w_size																							
1	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
2	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
3	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
4	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
5	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
6	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
7	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
8	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
9	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
10	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
15	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
20	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
25	30	40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
30		40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
35		40	50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
40			50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600		
60				50	60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600	
80					60	70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600	
100						70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600	
120							70	80	90	100	110	120	130	140	150	200	250	300	350	400	450	500	550	600

Figura 5.2: Combinaciones posibles que compusieron la grilla en su versión inicial

Generando un total de 408 casos para cada acción.

Por lo tanto, dado que se contaban con un total de 437 acciones, se llevaron a cabo 178.296 iteraciones y esto implicó aproximadamente una hora de procesamiento por acción.

En este caso, por ejemplo, para la acción BKNG, de cómo se ven los comportamientos de las distintas métricas en la grilla con distintos p y w para una misma acción.

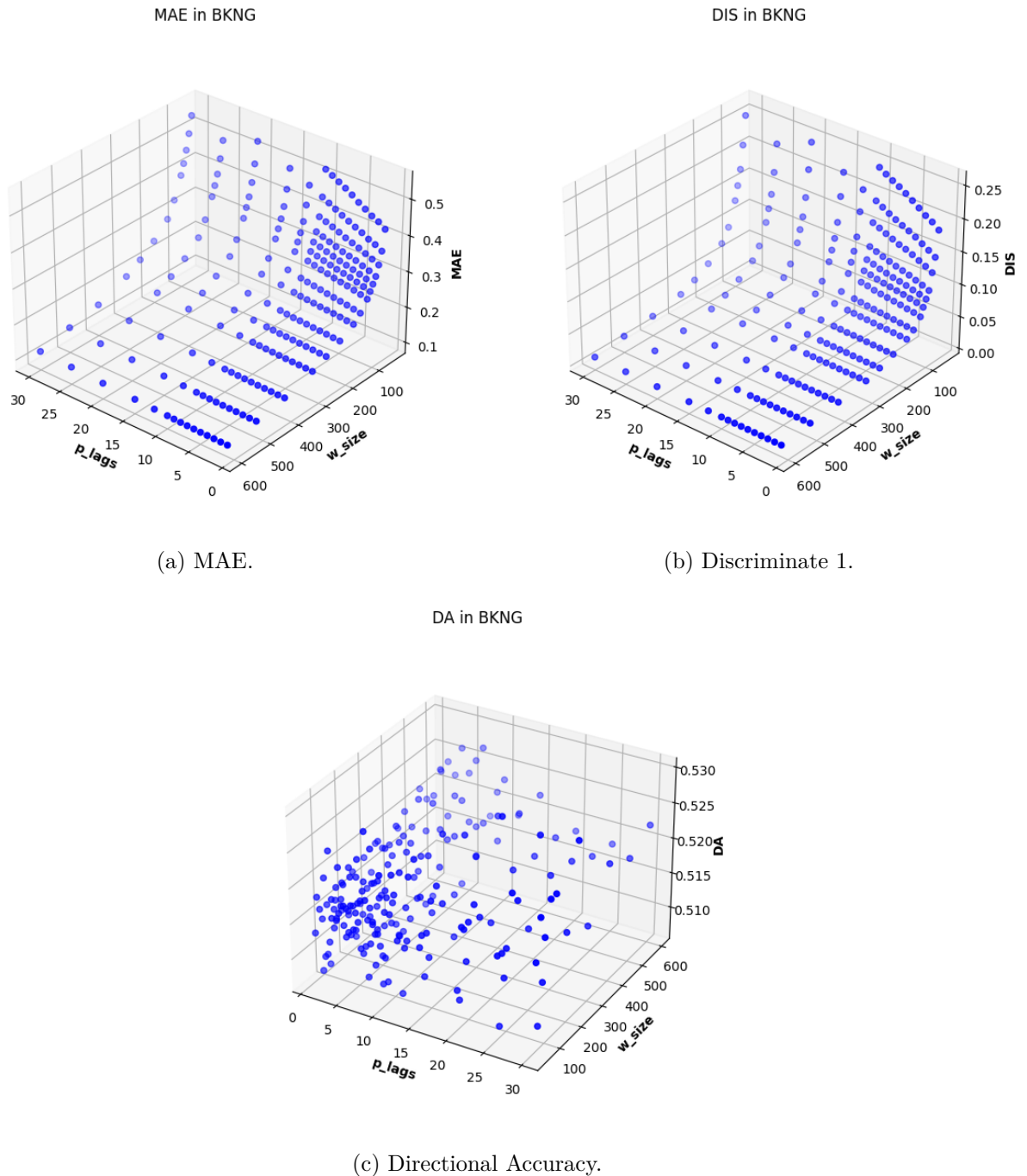


Figura 5.3: Métricas obtenidas de la grilla para cada combinación, activo BKNG

Gracias a esto, se logró obtener una vista general del comportamiento y rendimiento del estimador para cada métrica.

De los resultados de la grilla se pudo extraer que, para MAE y Discriminante 1, mientras menor sea el p mejores son sus resultados y en cuanto a w este tiene que ser más bien alto. En cambio, para el DA el gráfico resultante no permite extraer conclusiones similares a las otras métricas, debido a su comportamiento.

5.1.2. Robustez

Para comprobar que los resultados obtenidos fuesen correctos y robustos, indicando que los resultados fuesen aplicables a series de tiempo más reducidas o extensas, se optó por dividir el dataset en 3 conjuntos distintos:

- Set 1: conteniendo los primeros 1500 datos.
- Set 2: conteniendo los últimos 1500 datos.
- Set 3: conteniendo la totalidad de 3000 datos.

De tal modo que los experimentos aplicados realizaron en estos 3 conjuntos de datos.

5.2. Predictibilidad

5.2.1. Optimización de Métricas

Para tener una visión mas general del comportamiento de las métricas y las acciones, se buscaron los valores óptimos de cada acción en cuanto a; MAE, DIS y DA. Los resultados para el set 3 fueron:

1. MAE:

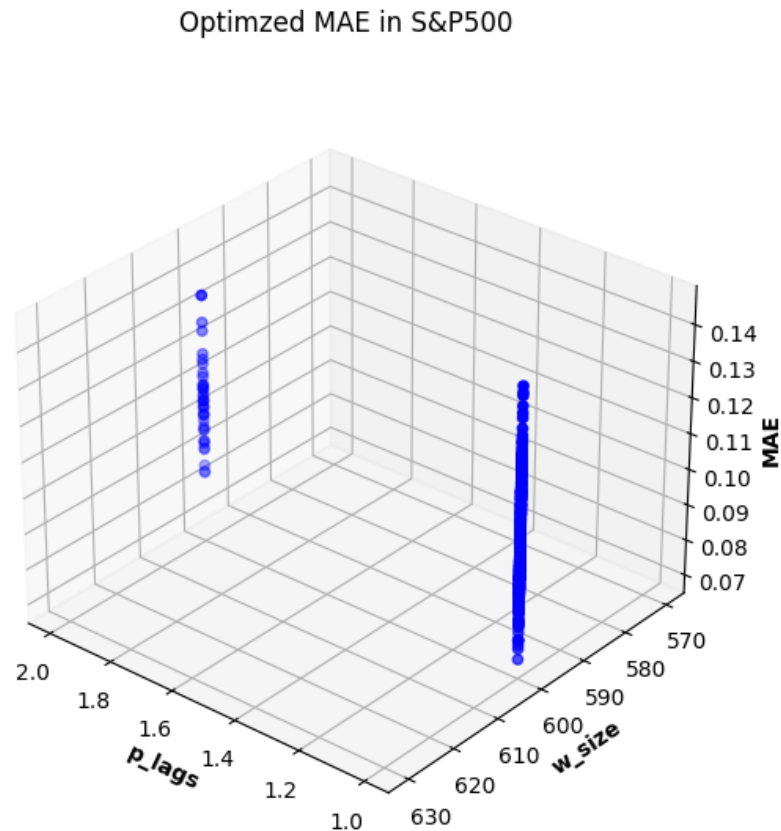


Figura 5.4: MAE optimizado para cada acción en al grilla

El análisis de la Figura 5.4 indica que, mientras mayor fuese w (tamaño de ventana) y más pequeño fuese p (rezagos), la métrica se optimizaba teniendo cada vez un valor menor; ubicándose todas las acciones con w igual a 600, por otro lado todas están entre p igual 1 o 2.

2. Discriminante 1

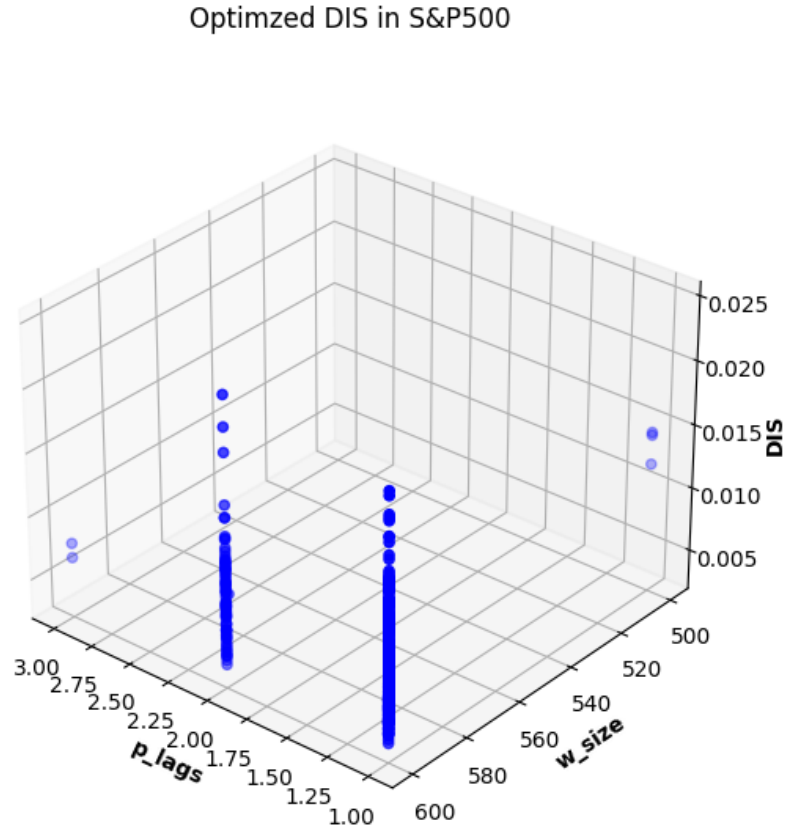


Figura 5.5: Discriminante 1 optimizado para cada acción en al grilla

La figura 5.5, donde se optimizó el DIS, tuvo un comportamiento muy similar que el MAE, teniendo un mejor rendimiento con w muy altos y p muy bajos. Las pequeñas diferencias recaen en que las mayoría de acciones optimizan DIS con w igual a 600, pero hay algunas que lo hacen con $w = 500$ y para p se distribuyen entre $p = [1, 2, 3]$, pero de todos modos siguen siendo los p mas bajos y los w mas altos existentes.

3. Direccional Accuracy

Optimized DA in S&P500

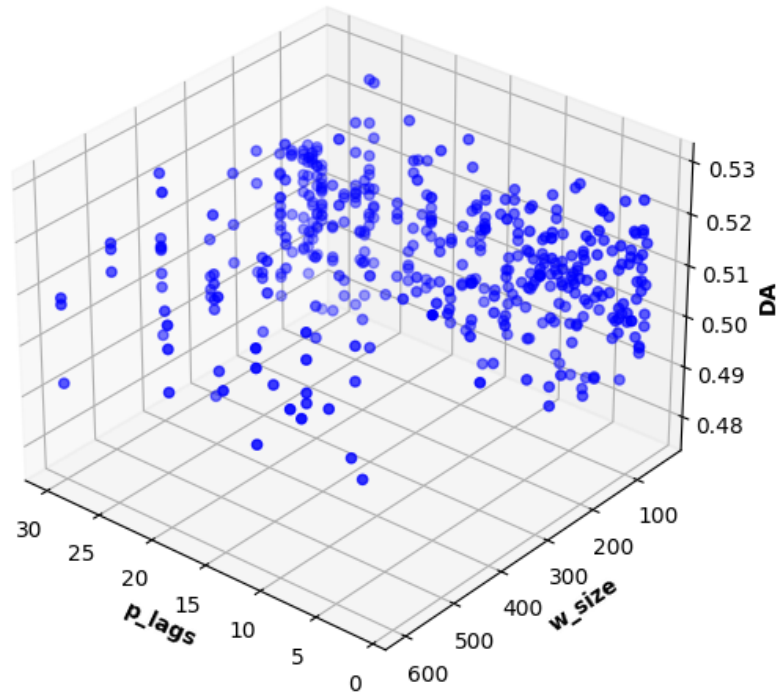


Figura 5.6: Directional Accuracy optimizado para cada acción en al grilla

A diferencia de las otras dos métricas, esta no presentó ningún comportamiento claro o evidente. Luego de un análisis más profundo, se obtuvo que:

- La mayoría de las acciones tuvieron predictibilidad siendo mayor a 0.5, es decir, 302 activos de 437 obtuvieron $DA > 0.5$, un 69% de las acciones.
- El 53% de las acciones, siendo estas específicamente 232, obtuvieron un w menor a 100.

Finalmente, se obtuvo que las métricas MAE y DIS se minimizan para su optimización, de tal modo que las acciones alcanzaban sus mínimos valores de DIS, MAE con los valores de w (ventanas) mas grandes y valores de p (rezagos) mas pequeños.

Si se ahonda en MAE, p solo puede tomar los valores de 1 o 2, lo cual es importante por el hecho de que esta métrica necesita como máximo 2 muestras de rezagos para mejorar su rendimiento, de modo que su predictibilidad y óptimo funcionamiento se basa en la menor cantidad de rezagos posibles y la ventana mas grande existente.

Algo similar ocurre con DIS, debido a que los p siguen siendo muy bajos sus posibles valores $p = [1, 2, 3]$ y en cuanto a w pertenece entre $[500, 600]$. Compartiendo comportamiento u características con MAE.

Por otro lado el DA parece ser que sus resultados son independientes respecto a cada acción, dado el comportamiento presentado, teniendo una amplia variedad de valores para p y w . Por lo que se apreció que el DA puede depender de otros factores para optimizarse, a diferencia del DIS o el MAE que tenían un claro comportamiento pra hallar sus mejores resultados.

Después de presentar en su totalidad los datos del set 3, se procede a la presentación del subconjunto del set 1, donde las métricas optimizadas fueron:

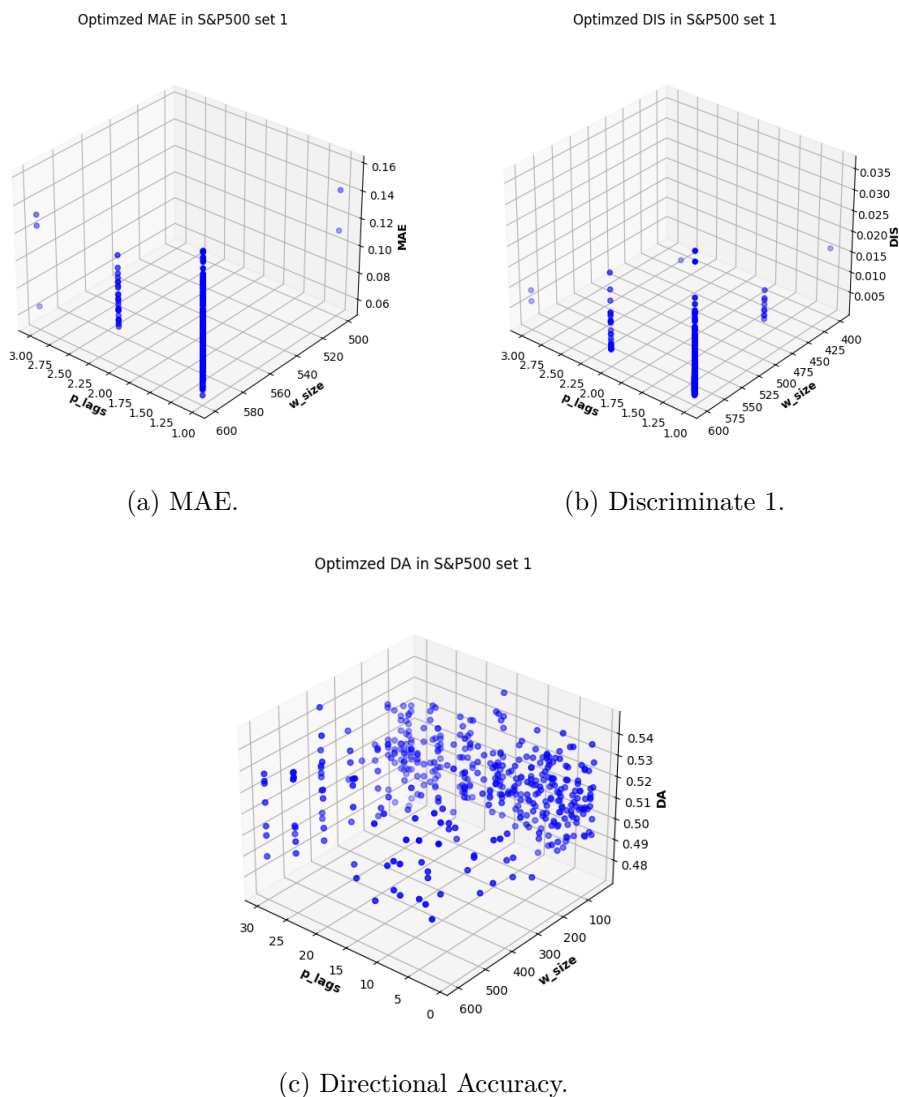


Figura 5.7: Acciones optimizadas para cada métrica para el set 1

- MAE: Posee los mejores valores con sus acciones teniendo w entre 500 y 600, en cambio para p entre 1,2 o 3.
- DIS: Se optimizó con un w entre 400, 500 o 600 y para los p desde 1,2 o 3.
- DA: teniendo un comportamiento mas disparejo del cual es difícil obtener conclusiones. Pero teniendo 326 acciones con un $DA > 0.5$ siendo un 75.5 % de todas las acciones.

Mientras que la optimización resultante del subconjunto set 2 fue:

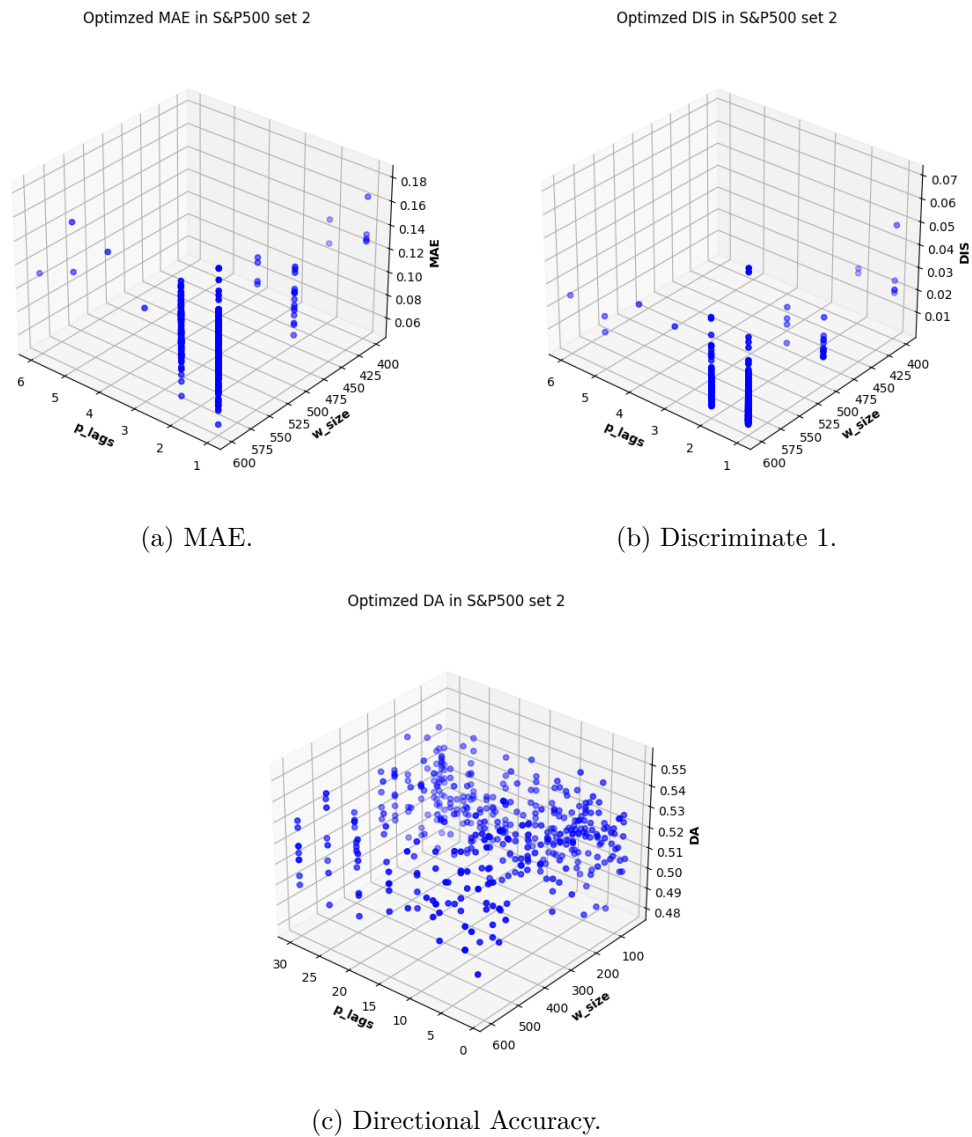


Figura 5.8: Acciones optimizadas para cada métrica para el set 2

- MAE: Posee los mejores valores con sus acciones teniendo w entre 400, 500 o 600, en cambio para p entre 1,2,3,4 o 5.
- DIS: Se optimizó con un w entre 300, 400, 500 o 600 y para los p desde 1 hasta el 10. Lo cual ha sido lo mas destacable en cuanto a diferencias entre el set 1 y el set 3, aunque sigue conservando una mayor densidad de resultados en los mismos valores que los otros 2 sets (p igual a 1 y w igual a 600).
- DA: teniendo un comportamiento igualmente disparejo al de los otros 2 sets, del cual es difícil obtener conclusiones. Pero con 388 acciones con un $DA > 0.5$, es decir, un 88.7% del total de las acciones

Como se observó, los comportamientos fueron similares entre los 3 sets, lo cual demostró robustez en el procedimiento realizado, aun así cabe destacar que en los sets con menores

datos, es decir, set 1 y 2 obtuvieron DA máximo mas alto que el set 3.

Otro punto importante fue que el set 2 presentó los niveles mayores de DA, llegando una acción a tener sobre el 55 %, la cual respeta todos los resultados obtenidos, debido a que esta tiene en promedio rachas mas largas y mayores variaciones de volumen.

5.2.2. Representación de señales

Con el fin de identificar la veracidad del DA y que no hubiesen fallos en su implementación, debido a su comportamiento.

Se decidió crear representaciones nuevas para la señal original y para la predicción. Donde se evaluaba la dirección de la señal, de tal modo que si su valor subía respecto al día anterior, se le asignaba un 1 en el día correspondiente, y si su valor bajaba respecto al día anterior, se le asignaba un valor de -1 en ese día.

Con esto se obtuvieron 2 nuevas señales entre 1 y -1 que representaban la dirección de la señal original y de la predicción. Esto se puede observar en la figura 5.9.

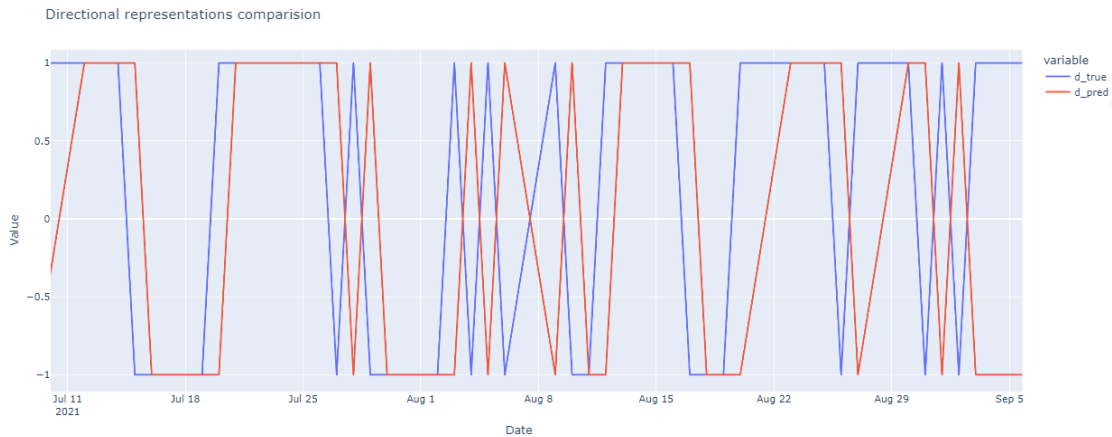
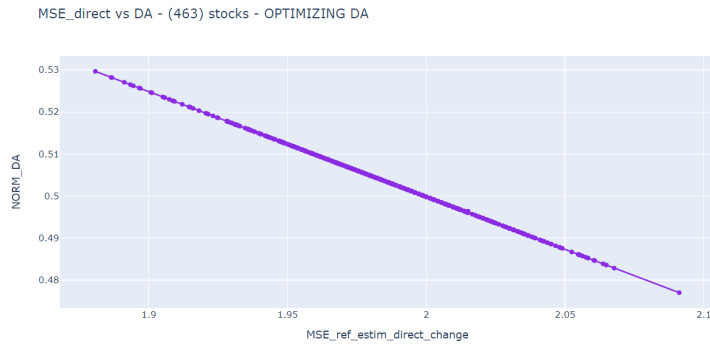
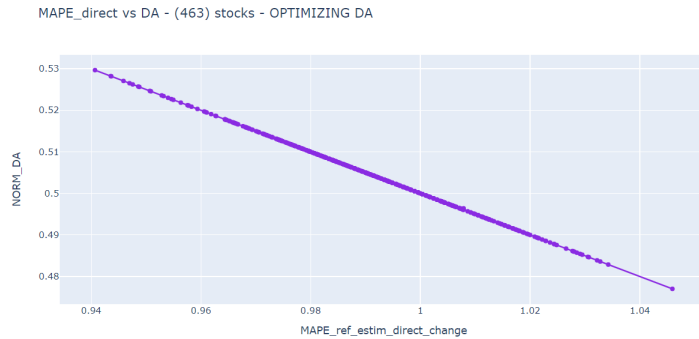


Figura 5.9: Representación de señales por su dirección

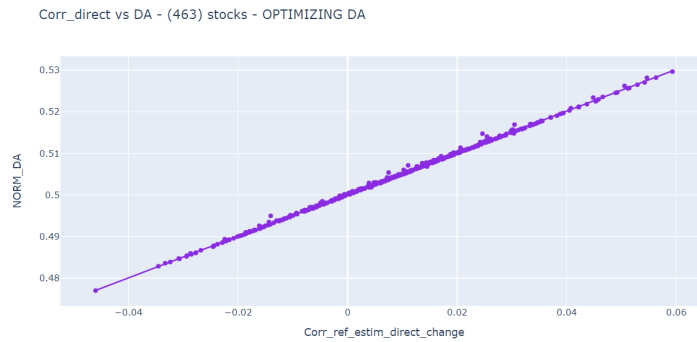
Posteriormente, se decidió aplicar 3 métricas a estas representaciones, como lo serían MSE, el MAPE y Correlación. Sus resultados fueron:



(a) MSE vs DA



(b) MAPE vs DA



(c) Correlación vs DA

Figura 5.10: Métricas obtenidas gracias a las representaciones, comparadas con el DA óptimo.

Como se puede ver en la figura 5.10, los resultados obtenidos tuvieron un comportamiento lógico con el Directional Accuracy. En las figuras 5.10 (a) y (b), se dedujo que existe una clara correlación negativa entre el MSE y el MAPE con el Directional Accuracy. Dicho de otra forma, se deduce que mientras menor sea el error (ya sea MAPE o MSE) entre las curvas direccionales, mayor será el Directional Accuracy.

Por otra parte, en la figura 5.10 (c) la correlación es positiva, concluyendo que mientras mayor correlación exista entre las señales direccionales, mayor será el DA, y que una correlación negativa entre ellas implicará un DA negativo o más bien menor a 0.5.

Esto significó que el Directional Accuracy se implementó de manera correcta, con lo cual

se precedió a seguir con más experimentos que aportaran en la explicación de su comportamiento y/o si existe posibilidad de extraer nuevas conclusiones.

El resultado que mas llamó la atención de esta sección fue la figura 5.2.2 (c), puesto que presentó valores de correlación negativos entre las señales direccionales, a su vez esto implicaba que tuviesen también un DA negativo. Por un DA negativo se sobre entiende que es un valor menor a 0.5, debido a que 0.5 es el peor resultado, significando que la predicción es nula, puesto que esta pasaría a ser aleatoria.

Adicionalmente el hecho de que una acción tenga valores menores a 0.5, implica que el algoritmo comienza a predecir la dirección contraria de la señal original, es decir, que se asemejaba mas a la señal original multiplicada por -1 , que a la original misma.

5.2.3. Relación entre Métricas

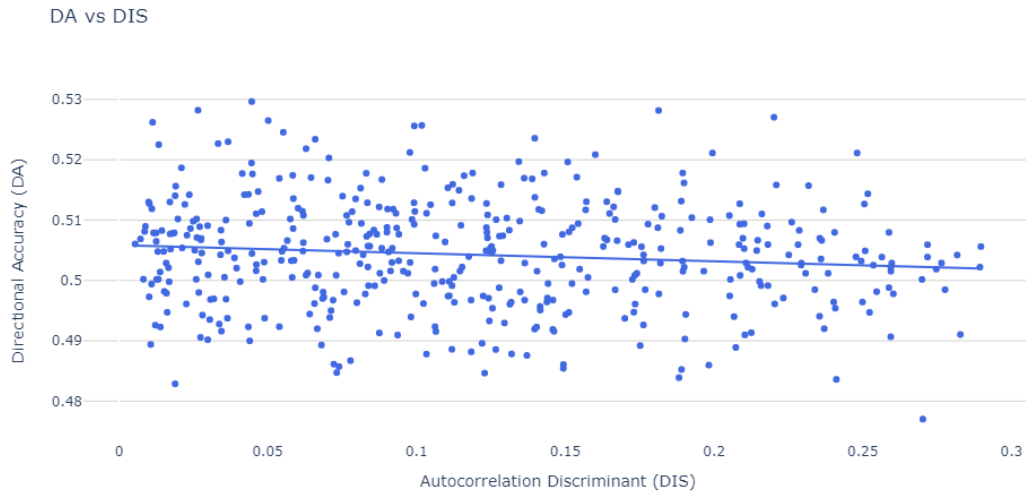


Figura 5.11: DA vs DIS

La relación que se mostró en la figura 5.11 es muy débil, indicando que mientras menor DIS, se obtendría un mejor DA, aun así, la pendiente de la regresión es demasiado pequeña, por lo que la cantidad de excepciones a esa relación es demasiado alta. Siendo lo mas aproximado decir que entre estas métricas prácticamente no poseen relación o que al menos no es determinante para extraer conclusiones al respecto sobre como una pueda afectar a la otra.

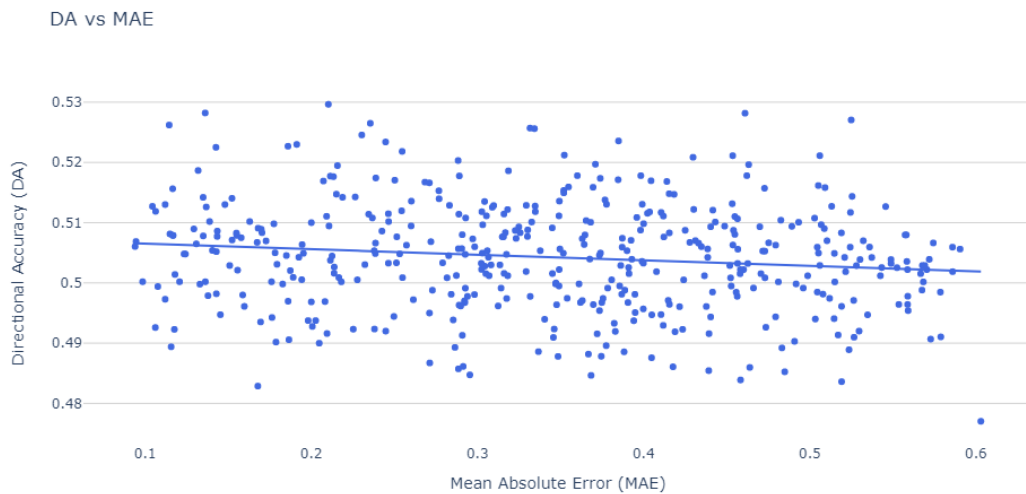


Figura 5.12: DA vs MAE

Al comparar MAE y DIS, sucede algo similar al caso anterior, teniendo el mismo comportamiento, con una relación lo suficientemente débil como para que una métrica influya sobre la otra.

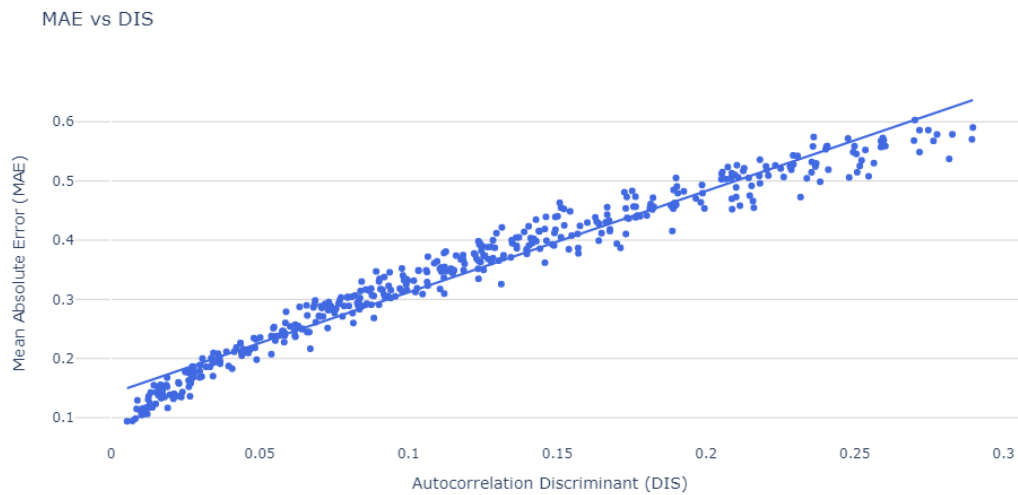


Figura 5.13: MAE vs DIS

MAE y DIS están correlacionadas positivamente como se pudo apreciar en la figura 5.13, es decir, si una aumenta la otra también y vice versa. Lo cual refleja lo similar que son las métricas. Sin embargo, estas no son directamente proporcionales a pesar de ser muy parecidas, con lo cual se asemejan a lo que están evaluando pero no son iguales.

Las relaciones que presentó el DA con DIS o con MAE, fueron muy débiles, como se apreció en las figuras 5.11 y 5.12. Pudiendo concluir de esto, que a pesar de que las leves correlaciones digan que a menor DIS o MAE se obtendrá un mayor DA, existen demasiados contraejemplos a esta conjetura, mostrando que no siempre un bajo MAE o DIS implican un

alto DA.

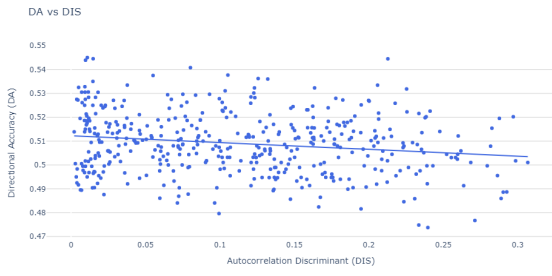
Como lo podrían ser el top 5 acciones según cada métrica optimizada:

Tabla 5.1: Top 5 acciones optimizadas para DA, MAE y DIS

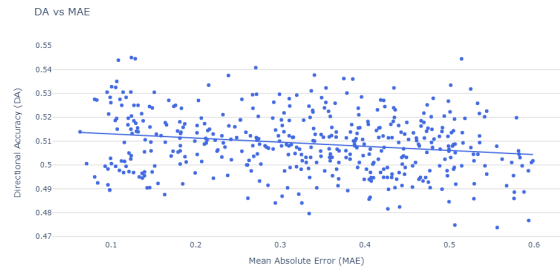
	Stock	w_size	p_lags	MAE	DA	DIS
DA Óptimo	BKNG	250	20	0.21	0.5293	0.0446
	IRM	500	25	0.136	0.5282	0.0265
	DLR	70	25	0.461	0.5281	0.1813
	INCY	60	25	0.524	0.527	0.2201
	PKI	150	15	0.235	0.5265	0.0501
MAE Óptimo	DXC	600	1	0.0702	0.4965	0.0099
	CTAS	600	1	0.0732	0.4681	0.0034
	CMG	600	1	0.0742	0.4823	0.0049
	NWL	600	1	0.0755	0.5115	0.0058
	HAL	600	1	0.0756	0.4940	0.0055
DIS Óptimo	CTAS	600	1	0.0732	0.4681	0.0034
	WST	600	1	0.0809	0.4848	0.0037
	MSCI	600	2	0.0850	0.4789	0.0037
	MKTX	600	1	0.0849	0.4831	0.0040
	ADBE	600	2	0.0867	0.4756	0.0041

Para los otros sets se obtuvo que:

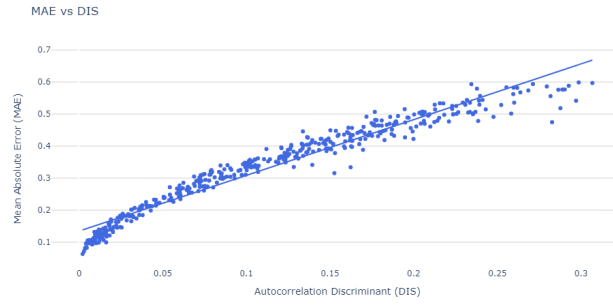
- SET 1:



(a) DA vs DIS



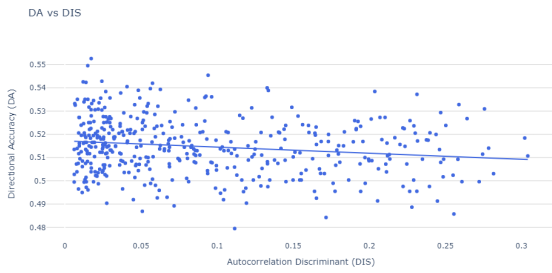
(b) DA vs MAE



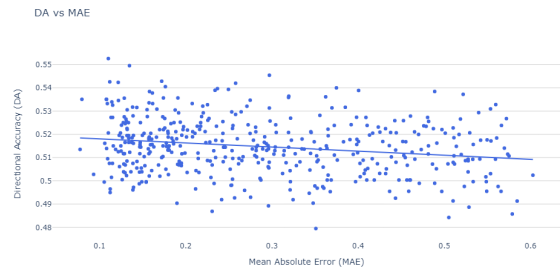
(c) MAE vs DIS

Figura 5.14: Relación entre Métricas Set 1.

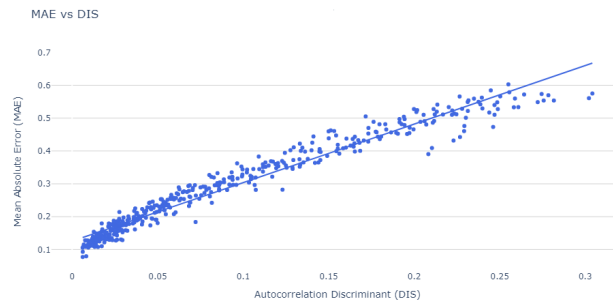
• SET 2:



(a) DA vs DIS



(b) DA vs MAE



(c) MAE vs DIS

Figura 5.15: Relación entre Métricas Set 2.

Donde el comportamiento y las correlaciones entre las métricas se mantiene entre el set 1, el set 2 como se apreció en las figuras 5.14, 5.15 y en el set 3 tal como se puede ver en la sección 5.2.3 .

5.3. Predictibilidad ARIMA

Otro punto para tener robustez en lo realizado corresponde a la implementación de ARIMA. Donde el experimento consistió en realizar un cambio a la hora de predecir con el algoritmo implementado, donde se decidió reemplazar LPC por ARIMA, con el fin de ver las diferencias a la hora de los resultados. Se realizó este experimento con las 10 mejores acciones en cuando a DA con LPC y también a las 10 peores.

Para LPC las 10 mejores acciones(mayor DA):

Tabla 5.2: Los 10 mejores valores de DA para LPC

stock	w_size	p_lags	MAE	DA	DIS
BKNG	250	20	0.210277	0.529647	0.044639
IRM	500	25	0.136151	0.528211	0.026576
DLR	70	25	0.461032	0.528167	0.181376
INCY	60	25	0.524933	0.52705	0.220155
PKI	150	15	0.235519	0.526501	0.050182
PENN	500	1	0.114482	0.52621	0.011326
ILMN	100	20	0.331644	0.525699	0.101788
RCL	90	15	0.334316	0.52561	0.099271
WTW	250	25	0.230383	0.524554	0.055252
BWA	50	1	0.384905	0.523567	0.139748

Por otro lado los resultados de estas 10 acciones en ARIMA:

Tabla 5.3: Resultados en ARIMA para las 10 mejores acciones

Stock	w_size	p_lags	MAE	DA	DIS
INCY	70	30	0.289371748	0.52919085	0.083691
PKI	150	30	0.188584232	0.525798526	0.033946
BKNG	400	30	0.13919676	0.525202001	0.021551
IRM	500	30	0.113047978	0.523809524	0.017578
DLR	60	30	0.309690412	0.522966996	0.097567
WTW	500	30	0.131036512	0.521408563	0.015796
PENN	400	30	0.121131064	0.521354367	0.013208
BWA	200	30	0.164089451	0.518399428	0.025334
RCL	125	30	0.207767507	0.512874043	0.040186
ILMN	80	30	0.236244864	0.509421035	0.066218

Con las 2 tablas a mano fue fácil notar que el ranking de las 10 mejores acciones obtenidas en LPC, referente a la metrica DA, cambió con ARIMA y que el mayor DA se obtuvo con el análisis LPC.

Luego de un análisis detallado de cada tabla se observó que todas las acciones mejoraron sus valores de MAE y DIS, es decir, disminuyeron sus errores mejorando en ambas métricas. Por lo que para estas 2 métricas ARIMA mejoró los resultados obtenidos a LPC.

En cambio para DA, se obtuvieron diversos resultados, debido a que para las acciones: BKNG, IRM, DLR, PKI, PENN, ILMN, RCL, WTW y BWA. ARIMA empeoró sus resultados, reduciendo su DA. Sin embargo para INCY mejoró su valor de DA en comparación al obtenido con LPC, siendo la única acción que mejoró sus 3 métricas con ARIMA.

Como se mencionó anteriormente LPC corresponde a una variante o sub caso de ARIMA con ciertas variaciones. Una de las diferencias mas importantes es que LPC no tiene componente de memoria, mas conocido como Moving Avarage(MA), el cual debe permitir la mejora en las métricas como el MAE o el DIS a la hora de comparar ARIMA con LPC, demostrando la importancia de este término en el modelo. A pesar de esto, por lo que se obtuvo en los resultados, parece no ser determinante a la hora de mejorar DA, presentando un aumento en esta métrica en tan solo en el 10 % de las acciones sometidas a prueba.

En cuanto a los 10 peores activos con DA mas bajo en LPC:

Tabla 5.4: Las 10 acciones con mas bajo DA

stock	w_size	p_lags	MAE	DA	DIS
NVDA	40	3	0.463898	0.485975	0.198259
AVGO	125	12	0.288565	0.485734	0.073958
AMP	80	25	0.439191	0.48544	0.149368
FISV	50	8	0.484931	0.485249	0.189095
SHW	150	30	0.29544	0.484731	0.073215
KEY	100	25	0.368386	0.48465	0.122898
HBAN	50	7	0.458333	0.483893	0.188226
BRKb	40	6	0.519151	0.483609	0.241101
LIN	400	25	0.167839	0.482878	0.018895
V	60	30	0.602973	0.477033	0.270185

Finalmente, las mismas 10 en ARIMA:

Tabla 5.5: 10 peores acciones en DA, corridas con ARIMA

Stock	w_size	p_lags	MAE	DA	DIS
NVDA	80	30	0.266730272	0.490578965	0.065929
FISV	90	30	0.267576474	0.482296322	0.066074
AVGO	200	30	0.193652059	0.481243301	0.033974
SHW	500	30	0.091685057	0.480192077	0.006329
BRKb	60	30	0.335393489	0.477373256	0.108496
AMP	100	30	0.233177508	0.476716109	0.047754
KEY	100	30	0.252231243	0.476371162	0.060927
HBAN	400	30	0.125044156	0.475952289	0.014763
LIN	200	30	0.210722936	0.474812433	0.040518
V	500	30	0.09819742	0.468987595	0.007053

Para las 10 peores acciones se obtuvo que; cambió el orden de las 10 peores entre LPC y ARIMA. Además de que el peor DA alcanzado fue por parte de ARIMA.

En este caso al igual que en el anterior, mejoró a todas las acciones en lo que respecta a las métricas del MAE y DIS. Pero en cuanto al DA, mejoró el valor de la acción NVDA, pero empeoró el valor de todas las otras acciones.

5.4. Simulación de trading

En la simulación de trading se buscó observar las acciones que generaron una mayor ganancia, para esto se realizó el backtesting para todos los sets optimizando cada métrica por separado.

5.4.1. Optimizando DA

- SET 1:

Tabla 5.6: Top 10 acciones del Set 1

stock	DA	MAE	DIS	Strategy	Sharp Ratio	Calmar Ratio	Return	Buy hold
TRMB	0.525	0.132	0.020	two side stoploss takeprofit reversed	1.144	3.602	510.461	7.930
WMB	0.544	0.109	0.010	two side stoploss takeprofit reversed	0.844	2.546	457.968	-21.336
ADSK	0.525	0.174	0.034	two side stoploss takeprofit reversed	1.175	3.089	532.767	88.307
STX	0.527	0.083	0.005	two side stoploss trailing reversed	1.018	3.466	277.778	-14.809
WDC	0.520	0.107	0.006	two side stoploss trailing reversed	0.755	1.274	158.338	9.596
APA	0.508	0.182	0.032	two side stoploss trailing reversed	0.382	0.417	73.545	-54.731
OXY	0.506	0.226	0.056	two side stoploss trailing reversed	0.599	0.674	109.660	-17.847
DRE	0.511	0.212	0.037	two side stoploss takeprofit reversed	1.024	1.147	235.997	108.740
AMD	0.520	0.435	0.163	two side stoploss trailing reversed	0.330	0.481	108.193	-5.721
PNR	0.527	0.499	0.200	two side stoploss trailing reversed	0.782	1.316	202.171	91.995

- SET 2:

Tabla 5.7: Top 10 mejores acciones para el set 2

stock	DA	MAE	DIS	Strategy	Sharp Ratio	Calmar Ratio	Return	Buy hold
PVH	0.522	0.202	0.041	two side stoploss trailing reversed	0.881	2.067	674.820	-27.388
CCL	0.513	0.179	0.030	two side stoploss takeprofit reversed	0.499	1.035	524.118	-85.387
PARA	0.535	0.108	0.013	two side stoploss trailing reversed	0.838	2.015	407.656	-51.843
VTR	0.529	0.258	0.059	two side stoploss takeprofit reversed	0.761	1.313	377.569	-23.183
TFX	0.512	0.184	0.025	two side stoploss takeprofit reversed	0.993	1.874	342.562	-0.675
EXPE	0.527	0.131	0.029	two side stoploss takeprofit reversed	0.591	1.344	252.405	-19.576
ILMN	0.536	0.320	0.093	two side stoploss trailing reversed	0.804	1.370	309.319	37.388
ALK	0.507	0.130	0.026	two side stoploss takeprofit reversed	0.467	0.898	168.383	-23.890
VNO	0.539	0.250	0.063	one side sell stoploss takeprofit	0.658	0.885	125.168	-66.956
DISH	0.521	0.319	0.107	two side stoploss trailing reversed	0.429	0.517	116.640	-67.043

• SET 3:

Tabla 5.8: Las 10 acciones con mejores retornos para el Set 3 optimizando DA.

Stock	DA	MAE	DIS	Strategy	Sharp Ratio	Calmar Ratio	Return %	Buy and Hold %
INCY	0.527	0.525	0.220	two side stoploss takeprofit reversed	0.669	0.869	3442.383	365.995
PARA	0.510	0.200	0.036	two side stoploss trailing reversed	0.525	0.492	486.929	3.702
IRM	0.528	0.136	0.027	two side stoploss trailing reversed	0.688	0.661	409.951	98.460
MHK	0.506	0.524	0.228	one side buy stoploss trailing	0.722	1.020	426.791	137.272
PVH	0.515	0.352	0.111	one side buy stoploss trailing	0.555	0.800	266.307	13.689
APA	0.509	0.425	0.152	two side stoploss takeprofit reversed	0.155	0.146	147.023	-72.985
VFC	0.513	0.104	0.010	two side stoploss trailing reversed	0.483	0.464	204.154	31.153
BBWI	0.521	0.352	0.098	one side buy stoploss trailing	0.502	0.375	237.429	72.921
EXPE	0.509	0.143	0.024	one side buy stoploss takeprofit	0.493	0.675	255.352	104.701
VNO	0.521	0.454	0.199	two side stoploss takeprofit reversed	0.210	0.170	83.772	-54.801

Donde existen 42 acciones que reportan beneficios por sobre la estrategia de Buy and Hold. Donde la distribución de las estrategias es:

Tabla 5.9: Estrategias en acciones que presentan mayor retorno sobre buy and hold

Strategies	SET 1	SET 2	SET 3
one side buy stoploss takeprofit	13	29	15
one side buy stoploss trailing	18	41	17
one side sell stoploss takeprofit	0	4	2
one side sell stoploss trailing	7	6	0
two side stoploss takeprofit reversed	10	21	3
two side stoploss trailing reversed	22	27	5
Total	70	128	42

El set 2 fue el set con mayor cantidad de acciones que logró tener ganancias por sobre la estrategia buy and hold, luego vino el set 1 y finalmente el set 3. Coincidiendo con el mismo orden sobre la cantidad de activos que tienen un $DA > 0.5$ en cada set.

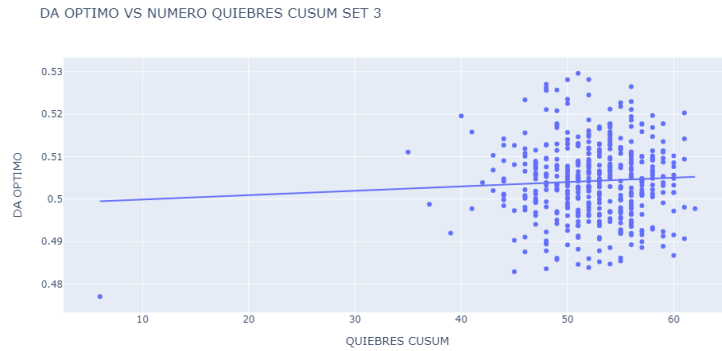
De las cosas más importantes que se pueden extraer de los resultados, es que no siempre un mayor DA implica una mayor ganancia a la hora de ganar dinero en trading, debido a que hay que considerar diversos aspectos, como lo serían las comisiones, el trayecto de la señal, volatilidad, riesgo, entre otros. Ya que si esta fluctúa demasiado a pesar de tener una gran predictibilidad, puede que en el camino pierda mas de lo que se gane. Por lo que este resultado se puede extrapolar a distintos tipos de predictores, con la enseñanza de que si bien la predictibilidad es un factor muy importante, no significa que este sea el único a la hora de buscar mayores retornos, por lo que hay que observar otros factores posteriormente a la obtención de predictibilidad, tales como; Calmar Ratio y Sharpe Ratio.

Los resultados de las simulaciones de trading optimizando las otras 2 métricas se pueden observar en el Anexo A y B.

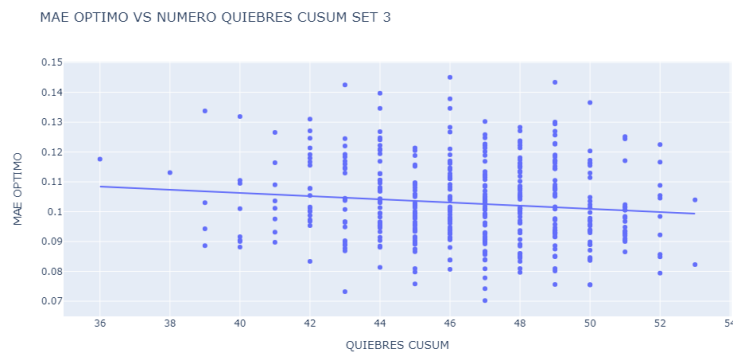
5.5. Determinantes de Predictibilidad

5.5.1. Análisis de Quiebres

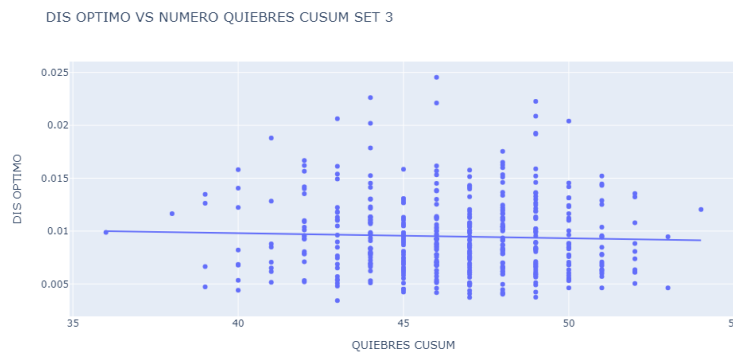
Se realizó la implementación de un detector de quiebres denominado CUSUM [34]. Con el fin de buscar si la cantidad de quiebres que tiene cada activo se correlaciona principalmente con el DA o alguna otra de las métricas.



(a) Para DA



(b) Para MAE



(c) Para DIS 1

Figura 5.16: Cantidad de quiebres contra las distintas métricas.

Para cada métrica se encontraron relaciones muy pequeñas. En la figura C.2(a) se aprecia una leve correlación positiva, de la cual se extrae, que mientras mayor cantidad de quiebres

mejor DA se podrá obtener.

En cuanto al MAE y al DIS, la relación es inversa, mientras mayor cantidad de quiebres, menor DIS o MAE se obtendrá.

Para los sets más pequeños, es decir, el 1 y el 2 con sus figuras en el Anexo C, se observó que las tendencias en su mayoría se mantuvieron, excepto por el DIS y el DA. Aun así las tendencias son muy débiles, con lo cual es entendible el cambio, debido que estos sets tienen la mitad de los datos en comparación al set 3.

5.5.2. Análisis de Tendencias

En este punto se decidió explorar si afectaba el hecho de que una señal fuese más constante en su crecimiento o decrecimiento, por lo que se realizó el siguiente experimento: Se define ΔX como la serie de tiempo realizada por su diferencia, es decir, $\Delta X(i) = X(i) - X(i - 1)$.

Esto permitió contar la duración de ciclos positivos y negativos, que es igual a contar cada racha de valores positivos o negativos consecutivamente, luego con los largos de cada racha se crearon histogramas para rachas positivas, negativas y rachas totales (negativas + positivas). Los histogramas fueron normalizados, por lo que sus valores representan la probabilidad de las rachas.

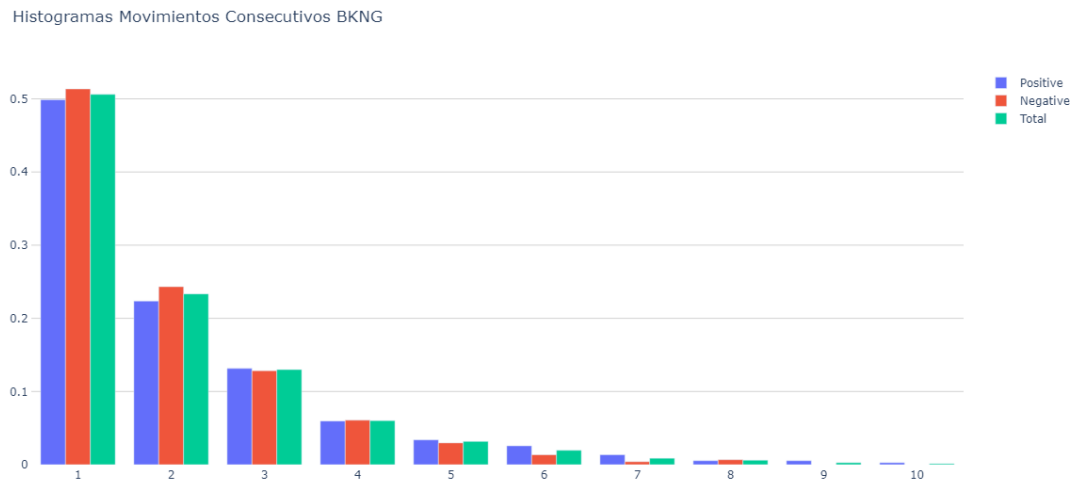


Figura 5.17: Ejemplo de histogramas para BKNG

Donde la distribución de los histogramas para todas las acciones fueron bastante similar a la que se mostró en la figura 5.17 correspondiente a BKNG. Donde predominantemente la racha que posee más frecuencia de ocurrencia es la racha de largo 1, la cual va decreciendo a medida avanza el largo de la racha.

Posteriormente se calcularon los histogramas para cada acción bursátil, donde se calculó el promedio de todos los histogramas. Y se extrajo lo siguiente:

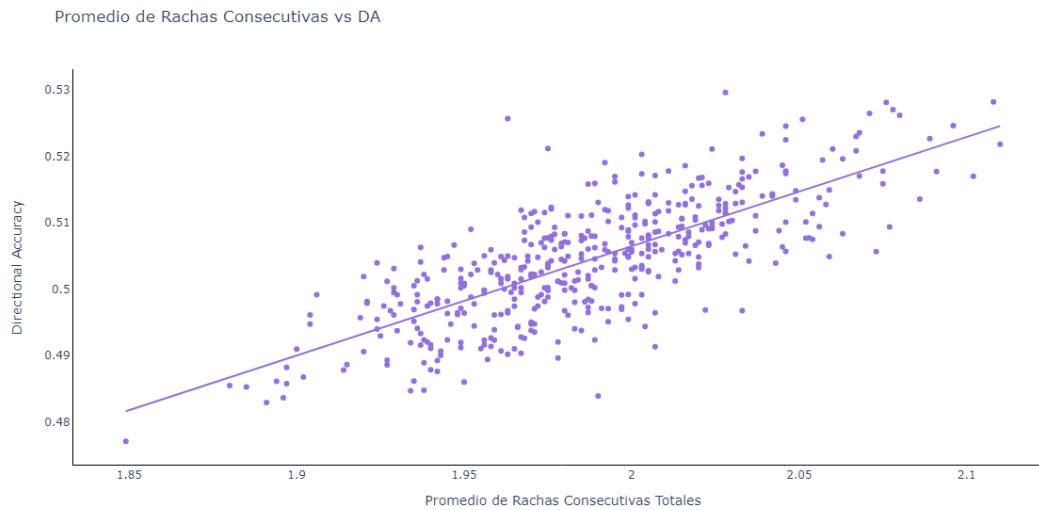
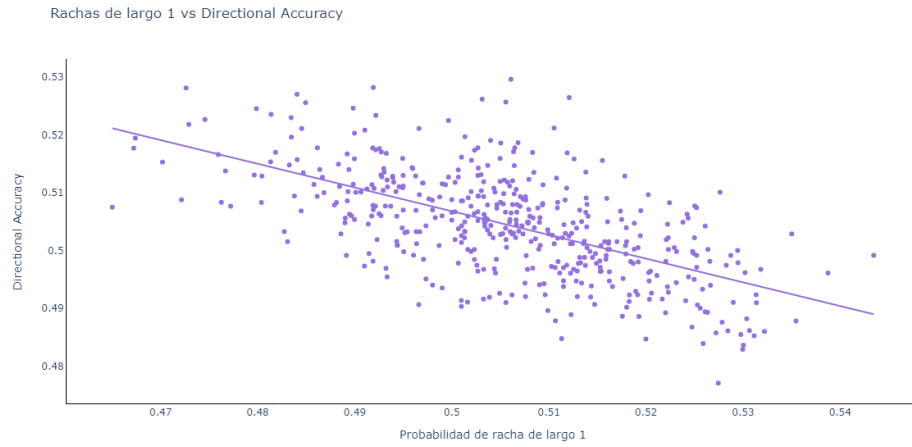


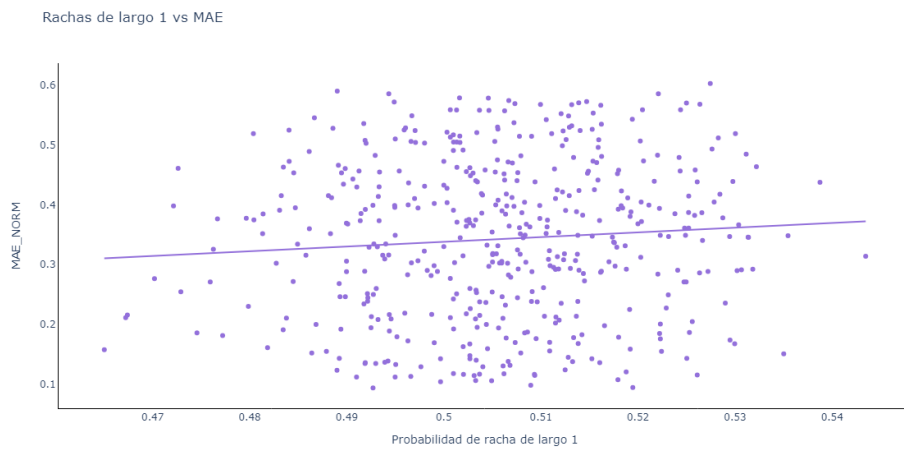
Figura 5.18: Promedio de rachas consecutivas para el S&P500

Con esto se observó que las acciones que tienen un promedio mas cercano a tener rachas de largo 1 más frecuentemente, tienen un peor DA, y que mientras más alto sea el promedio de las rachas, más DA se obtendrá.

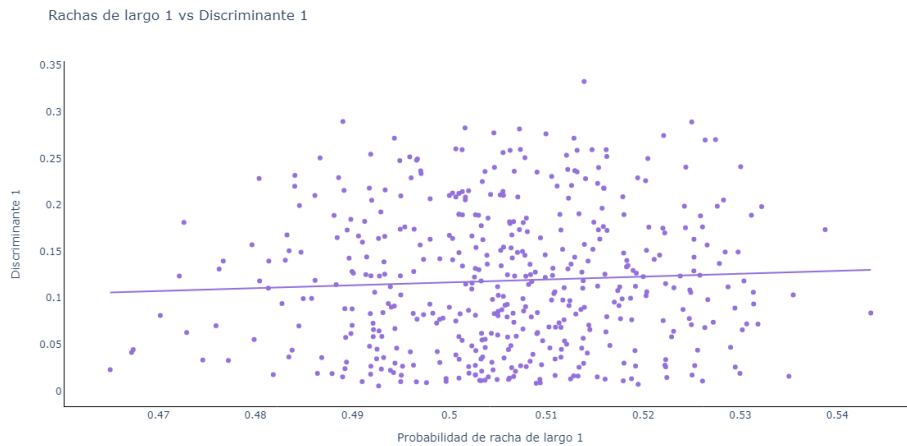
Otra forma de visualización, fue comparar la probabilidad de que cada activo tenga rachas de largo 1, y como esto se relacionaba con las métricas, para comprobar el resultado obtenido anteriormente.



(a) Para DA



(b) Para MAE



(c) Para DIS1

Figura 5.19: Rachas de largo 1 vs Métricas.

Con resultados para el DA se confirmó lo descubierto anteriormente, las acciones con mayor probabilidad de racha igual a 1 son las que poseen los peores niveles de DA y las con

menores probabilidad de racha igual a 1, es decir, las que tienen mayor probabilidad de tener rachas mayores a 1 son las que presentan un mayor rendimiento en cuanto a DA.

Finalmente se decidió comparar con las otras 2 métricas, donde se apreció que tenían leves tendencias de mejora, por lo que con menor probabilidad de rachas igual a 1, se obtendrían menores valores de MAE y DIS, en otras palabras, mejorarían levemente su rendimiento. En conclusión el mantener su estado, sea cual sea, subida o bajada, permite una mayor predictibilidad.

En lo que respecta los otros sets, los resultados corresponden a:

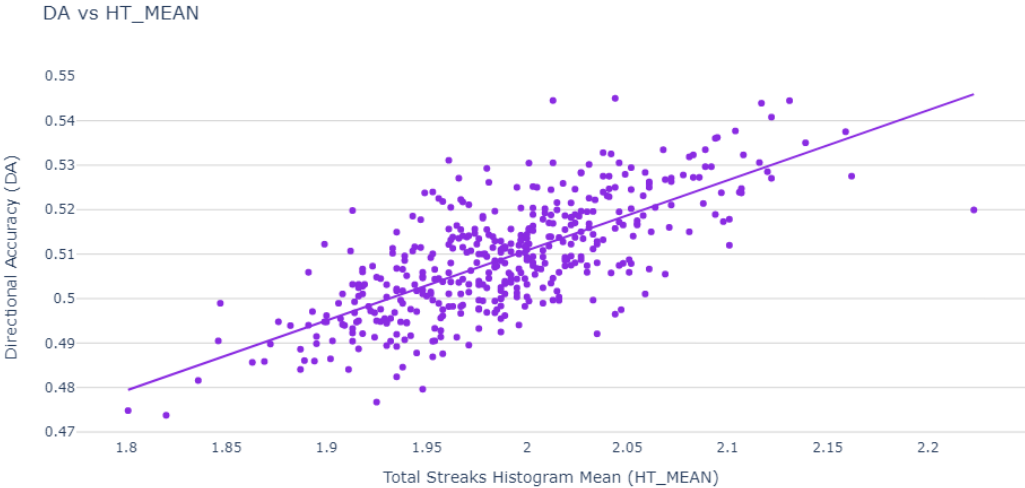


Figura 5.20: Promedio de rachas consecutivas para el S&P500 en el set 1

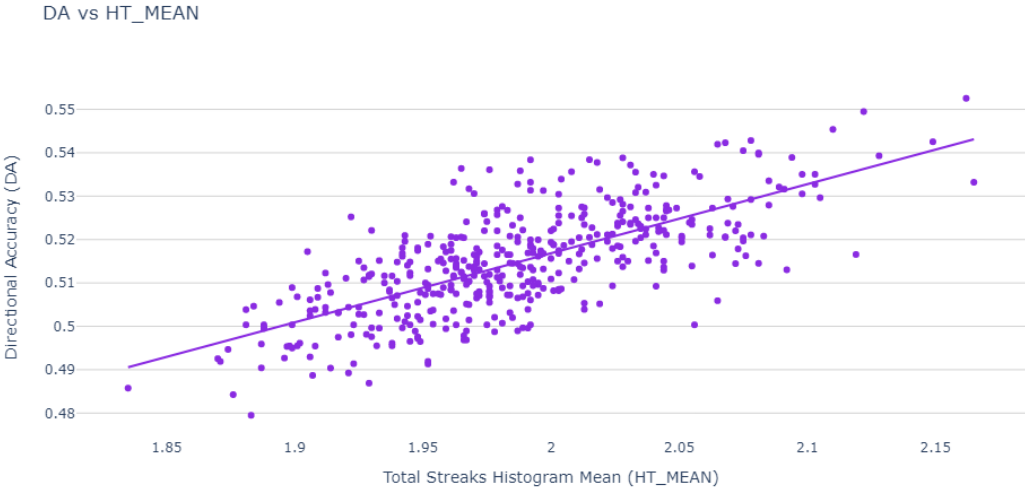


Figura 5.21: Promedio de rachas consecutivas para el S&P500 en el set 2

Para ambos sets, se conservó el mismo comportamiento con el set numero 3, indicando

que mientras mayor sea su racha promedio, mejor DA se podrá aspirar.

5.5.3. Análisis de Volumen

Para este experimento se quiso comprobar si el volumen afectaba el comportamiento del DA, por lo que en primer caso se comparó el volumen total transado de cada acción contra el DA, resultando en:

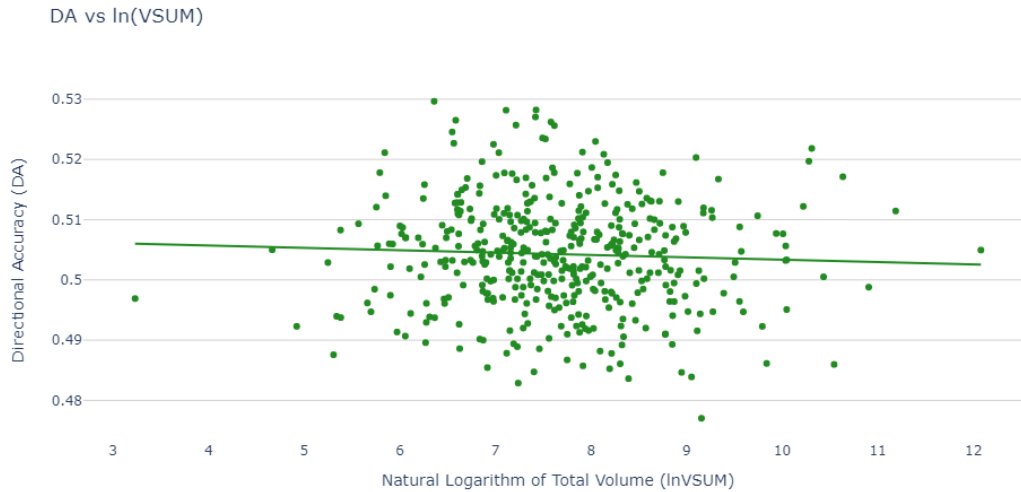


Figura 5.22: Logaritmo natural del volumen total vs DA, para cada acción

Demostró que existe una tendencia, pero que es demasiado débil y la distribución del gráfico no permite obtener conclusiones demasiado certeras, mas que a grandes rasgos, mientras mayor sea el volumen total transado, mas leve seria el DA.

Al no tener una incidencia muy clara el volumen total transado, se optó por analizar de otra manera el volumen, de tal modo, que se observen la variación de este mismo, buscando medir que tanto cambia el volumen transado a lo largo del tiempo.

Gracias a esto, se creó un coeficiente, el cual buscaba medir la variación del volumen, siendo definido como:

$$CV = \sigma_V / \mu_V$$

con σ_V la desviación del volumen y μ_V el promedio del volumen.

Los resultados obtenidos fueron:

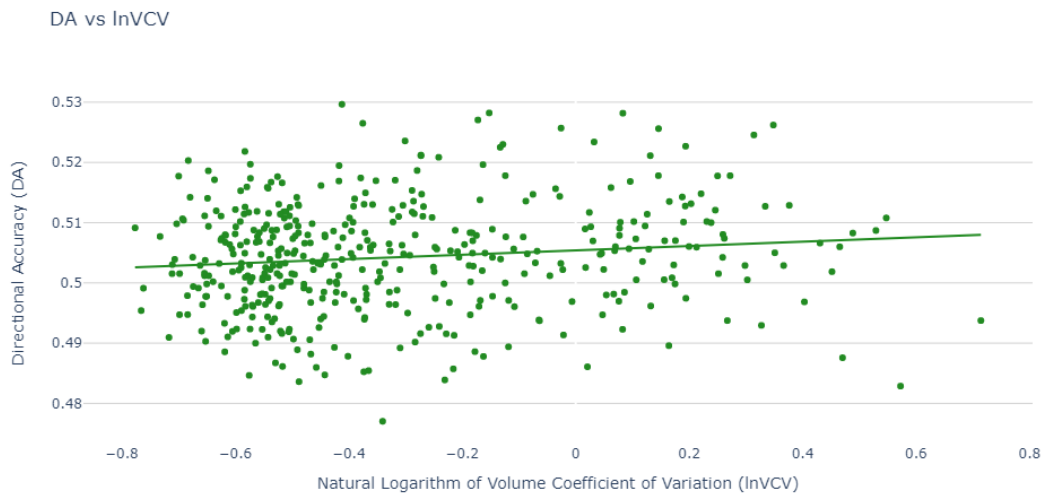


Figura 5.23: Logaritmo natural de la variación del volumen vs DA, para cada acción

Donde también se vislumbró una tendencia leve la cual describe que; mientras mas variación de volumen, mayor DA se podrá obtener.

En cuanto a los sets mas pequeños los resultados fueron:

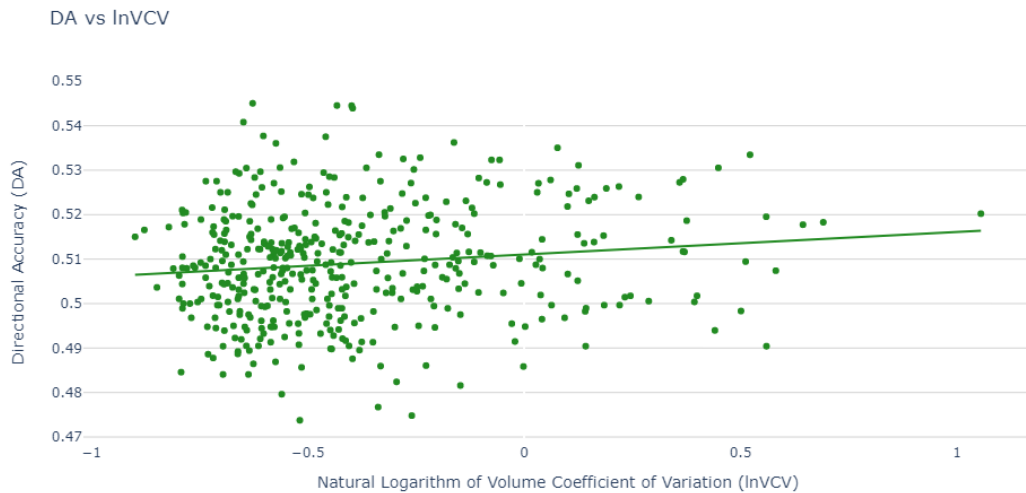


Figura 5.24: Logaritmo natural de la variación del volumen vs DA, para cada acción en el set 1

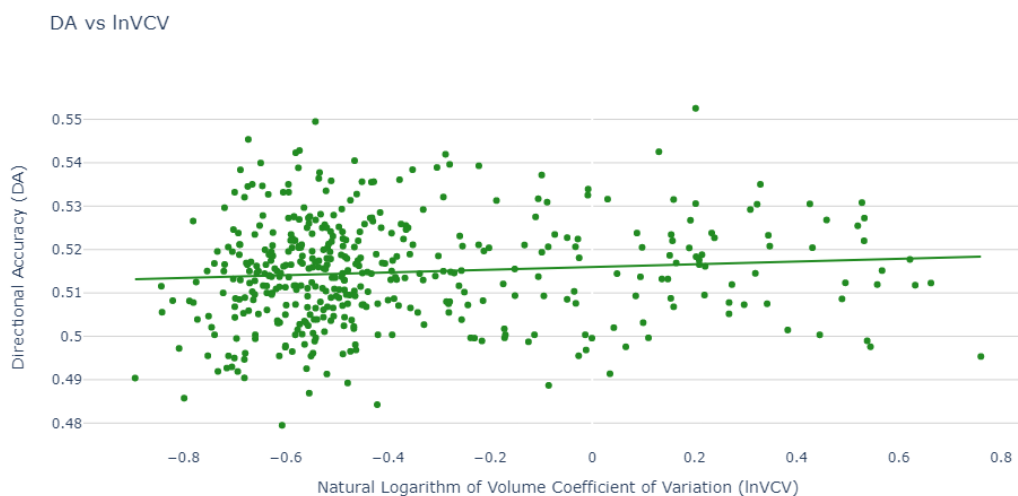


Figura 5.25: Logaritmo natural de la variación del volumen vs DA, para cada acción en el set 2

Nuevamente se respetaron los resultados obtenidos del set 3, obteniendo leves tendencias, indicando en ambos sets (set 1 y set 2) que a mayor variación de volumen se podrá tener un mayor DA.

5.5.4. Regresiones Multivariantes

El experimento consistió en realizar distintas regresiones lineales múltiples, cada una de estas basadas en diferentes combinaciones de las variables de la tabla 4.1, con el fin de buscar predecir el DA, Obteniendo así R^2 y $R^2_{Ajustado}$, en resumen, cuanto se ajusta el modelo a la variable deseada. Mostrando el mejor modelo para cada set.

Los mejores resultados obtenidos de cada set fueron:

5.5.5. SET 1

Tabla 5.10: Regresión lineal múltiple para explicar el DA en el set 1

Variables	Coef	std err	t	P> t	[0.025	0.975]
CONST	14.544	1.524	9.542	0	11.548	17.54
SPREAD_BETA	2.308	0.898	2.57	0.011	0.543	4.073
MARKET_BETA	-0.375	0.143	-2.626	0.009	-0.656	-0.094
DA_W_SIZE	0.000	0	1.667	0.096	-7.61E-05	0.001
DA_P_LAGS	0.032	0.004	7.422	0	0.024	0.04
PBPERSHARE_RMEAN	0.002	0.001	2.11	0.035	0	0.005
STREAK_HMEAN	17.843	0.773	23.078	0	16.323	19.363
ANALYST_DISPERSION_RMEAN	0.004	0.002	1.811	0.071	0	0.008
PUTCALLRATIO_RMEAN	0.017	0.03	0.571	0.568	-0.041	0.075
BIDASK_RMEAN	1.423	1.513	0.94	0.348	-1.551	4.396
HURST_RANDWALK_DAYS	0.001	0	2.672	0.008	0	0.001

Tabla 5.11: Resultados del modelo para el set 1

R-squared	0.62
Adj. R-squared	0.611
AIC	1063
Durbin-Watson	2.185
Breusch-Pagan	0.17
Shapiro-Wilk	0.381

5.5.6. SET 2

Tabla 5.12: Regresión lineal múltiple para explicar el DA en el set 2

Variables	Coef	std err	t	P> t	[0.025	0.975]
const	14.3288	1.891	7.579	0	10.612	18.045
SPREAD_BETA	-1.3063	0.667	-1.958	0.051	-2.618	0.005
MARKET_BETA	-0.3792	0.167	-2.273	0.024	-0.707	-0.051
DA_W_SIZE	0.0014	0	6.061	0	1.00E-03	0.002
DA_P_LAGS	0.0365	0.004	8.273	0	0.028	0.045
STREAK_HMEAN	17.8173	0.842	21.152	0	16.161	19.473
ANALYST_NUMBER_RMEAN	-0.0099	0.007	-1.502	0.134	-0.023	0.003
PUTCALLRATIO_RMEAN	0.0344	0.045	0.759	0.448	-0.055	0.123
BIDASK_RMEAN	0.1868	0.476	0.393	0.695	-0.748	1.122
PRICE_CUSUM_BREAKS	0.0357	0.019	1.872	0.062	-0.002	0.073

Tabla 5.13: Resultados del modelo para el set 2

R-squared	0.603
Adj. R-squared	0.594
AIC	1088
Durbin-Watson	2.012
breusch-pagan	0.369
shapiro-wilk	0.636

5.5.7. SET 3

Tabla 5.14: Regresión lineal múltiple para explicar el DA en el set 3

Variables	Coef	std err	t	P> t	[0.025	0.975]
const	12.6667	1.48	8.559	0	9.758	15.576
SPREAD_BETA	1.7066	1.238	1.379	0.169	-0.726	4.14
R_MSTD_LT	-0.1673	0.071	-2.34	0.02	-0.308	-0.027
DA_P_LAGS	0.0258	0.003	8.717	0	2.00E-02	0.032
STREAK_HMEAN	18.8536	0.749	25.172	0	17.381	20.326
PUTCALLRATIO_RMEAN	0.0113	0.026	0.43	0.667	-0.04	0.063
BIDASK_RMEAN	1.7263	0.503	3.43	0.001	0.737	2.716

Tabla 5.15: Resultados del modelo para el set 3

R-squared	0.617
Adj. R-squared	0.611
AIC	782.2
Durbin-Watson	1.927
breusch-pagan	0.148
shapiro-wilk	0.136

Comparando los modelos entre sets, se obtuvo que la regresión que mas se adapto al DA, fue la del set 1, llegando a tener los mayores niveles de R^2 y R_{ADJ}^2 . A pesar de todo los 3 sets poseen resultados muy similares, repitiendose las siguientes variables en los 3 casos:

- Spread Beta
- DA p Lags
- Streak Mean
- Bidask Rolling Mean
- Put call Ratio Rolling Mean

Donde las que presentan un mayor grado de importancia son: Spread Beta, DA p Lags y Streak mean.

Los resultados obtenidos fueron positivos, puesto que mostró que ciertas variables lograron explicar parte del comportamiento de la métrica DA, estas pudieron explicar alrededor de un 60 % del DA. Donde existen dos posibilidades sobre estos resultados; primero es buscar mas variables, de tal modo que los modelos puedan llegar a tener mejores R^2 y $Adj.R^2$; segundo, simplemente mediante esta técnica no se lograría explicar completamente su comportamiento debido a sus limitaciones.

Capítulo 6

Conclusión

Con el trabajo realizado, se concluyó que el análisis LPC es aplicable al mundo financiero, obteniendo capacidad predictiva en diversas acciones del S&P500. Observando entre las 3 principales métricas, existieron 2 que se centraron en el error y una orientada al acierto de las direcciones. A la hora de avanzar con los experimentos, se decidió hacer énfasis en acertar las direcciones, tomando como métrica principal el DA, debido a la importancia de esta en su uso primordial, siendo utilizada a la hora de invertir en activos; puesto que optimizar las métricas MAE o DIS no significó una mejora en el DA.

Logrando así realizar un posterior análisis y serie de experimentos que pudieron ayudar a encontrar diversas causas sobre la diferencia de predictibilidad entre distintas acciones. Como lo pudieron ser las rachas en los histogramas correlacionándose altamente con el DA, exponiendo que independientemente el signo de la racha, mientras más alto fuese el promedio de estas dentro de una acción, mayor DA se obtendría. Otro experimento fue la variación del volumen, que a pesar de ser levemente correlacionada, mostró que a mayor variación de volumen, mayor predictibilidad en cuanto a DA se podría tener.

El análisis de determinantes cumplió con el rol de ser una especie de mapeo de predictibilidad, que buscó justificar el comportamiento del DA con distintas variables, donde se coincidió con el punto anterior, siendo una de las variables más importantes; el análisis de tendencias, es decir, las rachas de los histogramas, a la hora de explicar el DA. Mostrando que al mantener su estado, el DA se beneficia; en cambio, al variar permanentemente perjudican la métrica.

Por otro lado, el tener un mayor DA no implicó tener más ganancias económicas en trading en todos los casos, si bien hay una leve tendencia hay que tener precaución, debido a que es muy importante tener en cuenta el cómo esta llegó hasta su precio actual, debido a que al ser muy riesgosa o volátil, esta puede reportar mayores pérdidas, sin importar su DA. Por lo cual se debe observar su comportamiento a través del tiempo, siendo de gran ayuda el Sharpe Ratio y el Calmar Ratio a la hora de elegir entre que acciones invertir.

Luego de la obtención de todos los resultados, se observó que claramente el predictor basado en LPC no puede competir de igual a igual con las técnicas más avanzadas de predicción, debido a que el estado del arte, con redes neuronales está más avanzado en este ámbito, logrando mayor predictibilidad como se pudo observar en la revisión literaria. Sin embargo,

esto no impidió realizar el estudio y demostrar la existencia de un nuevo método de predicción, que si bien no es el más avanzado permite entregar resultados coherentes encontrando capacidad predictiva en el área financiera, teniendo un coste computacional menor al que implican las redes neuronales.

Los objetivos planteados en un inicio se cumplieron en su totalidad logrando la implementación desde cero hasta su versión final de algoritmo de predicción basado en LPC. Obteniendo predicciones, realizando su respectivo análisis, comparándolos con ARIMA y finalmente realizando un mapeo o test de predictibilidad, el cual buscó encontrar el comportamiento de las acciones mediante sus determinantes.

A pesar de las dificultades de la investigación, como lo fueron diversos experimentos realizados sin resultados o conclusiones provechosas, los largos tiempos de cómputo para obtener los resultados, donde cometer un error era muy costoso en cuanto a los días que se perdían debido a la gran cantidad de datos siendo procesados en simultáneo, esto enseñó a ser muy metódico y cuidadoso a la hora de realizar cambios y experimentos nuevos. Cada paso fue necesario para la posterior creación de los experimentos que sí fueron concluyentes e importantes a la hora de obtener resultados. Da cuenta de lo importante que es ir intentando y descartando hipótesis para encontrar y formar un camino en la investigación.

Finalmente como trabajo futuro, se podría continuar:

- Aplicar ARIMA para todas las acciones, así comparar cada una con LPC para tener una generalización total al comparar ambas técnicas.
- Analizar la existencia de acciones con DA menor a 0.5, el porqué los modelos comienzan a predecir la dirección contraria a la señal original, cual serían los indicios o la justificación de este suceso.
- Realizar un test con redes neuronales sobre la predictibilidad, basándose o tomando de referencia al análisis de determinantes, pudiendo tomar las variables de mayor importancia obtenidas en este, de tal manera que la red neuronal las reciba como entrada y pueda decir si esa acción en cuestión tiene una alta o baja predictibilidad, sin necesidad de realizar todo el proceso de predicción de esa acción.
- Extrapolar el estudio a otros casos, como lo sería la renta fija u otros mercados, como lo podría ser el IPSA o cualquier mercado que sea de interés para los investigadores. Además de variar el periodo de los datos tomando más muestras o menos.

Bibliografía

- [1] Salemi Mottaghi, M., H. C. M., “A deep comprehensive model for stock price prediction,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, p. 11385–11395, 2023, doi:<https://doi-org.uchile.idm.oclc.org/10.1007/s12652-023-04653-2>.
- [2] Abu-Mostafa, Y.S., A. A., “Introduction to financial forecasting.,” *Applied Intelligence*, vol. 6, p. 205–213, 1996, doi:<https://doi.org/10.1007/BF00126626>.
- [3] “Progress and prospects of data-driven stock price forecasting research,” *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 100–108, 2023, doi:<https://doi.org/10.1016/j.ijcce.2023.03.001>.
- [4] Zhang, J., Cui, S., Xu, Y., Li, Q., y Li, T., “A novel data-driven stock price trend prediction system,” *Expert Systems with Applications*, vol. 97, pp. 60–69, 2018, doi:<https://doi.org/10.1016/j.eswa.2017.12.026>.
- [5] Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T., y Ngo, D. C. L., “Text mining for market prediction: A systematic review,” *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014, doi:<https://doi.org/10.1016/j.eswa.2014.06.009>.
- [6] Kumbure, M. M., Lohrmann, C., Luukka, P., y Porras, J., “Machine learning techniques and data for stock market forecasting: A literature review,” *Expert Systems with Applications*, vol. 197, pp. 1–15, 2022, doi:<https://doi.org/10.1016/j.eswa.2022.116659>.
- [7] Zhang, J., Cui, S., Xu, Y., Li, Q., y Li, T., “A novel data-driven stock price trend prediction system,” *Expert Systems with Applications*, vol. 97, pp. 60–69, 2018, doi:<https://doi.org/10.1016/j.eswa.2017.12.026>.
- [8] Rodríguez-González, A., Ángel García-Crespo, Colomo-Palacios, R., Guldrís Iglesias, F., y Gómez-Berbís, J. M., “Cast: Using neural networks to improve trading systems based on technical analysis by means of the rsi financial indicator,” *Expert Systems with Applications*, vol. 38, no. 9, pp. 11489–11500, 2011, doi:<https://doi.org/10.1016/j.eswa.2011.03.023>.
- [9] Fan Zhang, Y. D. y Liao, Y., “Financial data collection based on big data intelligent processing,” *International Journal of Information Technologies and Systems Approach*, vol. 36, pp. 1–13, 2023, doi:[10.4018/IJITSA.320514](https://doi.org/10.4018/IJITSA.320514).
- [10] Jiang, W., “Applications of deep learning in stock market prediction: Recent progress,” *Expert Systems with Applications*, vol. 184, p. 115537, 2021, doi:<https://doi.org/10.1016/j.eswa.2021.115537>.
- [11] Subakkar, A., G. J. S. J. A. L. G. J. Y. S. C., “An analysis on tesla’s stock price forecasting using arima model,” *Proceedings of the International Conference on Cognitive and Intelligent Computing*, vol. 2, pp. 83–89, 2023, doi:[10.1007/978-981-19-2358-6_8](https://doi.org/10.1007/978-981-19-2358-6_8).

- [12] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., y Oliveira, A. L., “Computational intelligence and financial markets: A survey and future directions,” *Expert Systems with Applications*, vol. 55, pp. 194–211, 2016, doi:<https://doi.org/10.1016/j.eswa.2016.02.006>.
- [13] Bustos, O. y Pomares-Quimbaya, A., “Stock market movement forecast: A systematic review,” *Expert Systems with Applications*, vol. 156, pp. 1–15, 2020, doi:<https://doi.org/10.1016/j.eswa.2020.113464>.
- [14] Xu, J., D. M. . d., “New robust lpc-based method for time-resolved morphology of high-noise multiple frequency signals,” *Irish Signals and Systems Conference (ISSC)*, vol. 31, pp. 1–6, 2020, doi:[10.1109/ISSC49989.2020.9180212](https://doi.org/10.1109/ISSC49989.2020.9180212).
- [15] Roy, S. K., Nicolson, A., y Paliwal, K. K., “Deep-lpc-mhanet: Multi-head self-attention for augmented kalman filter-based speech enhancement,” *IEEE Access*, vol. 9, pp. 70516–70530, 2021, doi:[10.1109/ACCESS.2021.3077281](https://doi.org/10.1109/ACCESS.2021.3077281).
- [16] André, P., Walton, P., y Yang, D., “Voluntary adoption of ifrs: A study of determinants for uk unlisted firms,” *SSRN Electronic Journal*, 2012, doi:[10.2139/ssrn.1978986](https://doi.org/10.2139/ssrn.1978986).
- [17] Campbell, J. y Thomson, S., “Predicting excess stock returns out of sample: Can anything beat the historical average?,” *Investopedia*, 2008, https://dash.harvard.edu/bitstream/handle/1/2622619/Campbell_Predicting.pdf?sequence=2.
- [18] Fama, E. y French, K., “The cross-section of expected stock returns,” 1992, <https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1992.tb04398.x>.
- [19] Brock, W., Lakonishok, J., y LeBaron, B., “Simple technical trading rules and the stochastic properties of stock returns,” 1992.
- [20] BAKER, M. y WURGLER, J., “Investor sentiment and the cross-section of stock returns,” *The Journal of Finance*, vol. 61, no. 4, pp. 1645–1680, 2006, doi:<https://doi.org/10.1111/j.1540-6261.2006.00885.x>.
- [21] BOLLERSLEV, T., “Generalized autoregressive conditional heteroskedasticity,” 1986, http://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdf.
- [22] Amihud, Y., “Illiquidity and stock returns: cross-section and time-series effects,” *Journal of Financial Markets*, vol. 5, no. 1, pp. 31–56, 2002, doi:[https://doi.org/10.1016/S1386-4181\(01\)00024-6](https://doi.org/10.1016/S1386-4181(01)00024-6).
- [23] Barberis, N. y Thaler, R., “Chapter 18 a survey of behavioral finance,” en *Financial Markets and Asset Pricing*, vol. 1 de *Handbook of the Economics of Finance*, pp. 1053–1128, Elsevier, 2003, doi:[https://doi.org/10.1016/S1574-0102\(03\)01027-6](https://doi.org/10.1016/S1574-0102(03)01027-6).
- [24] Vanstone, B. y Finnie, G., “An empirical methodology for developing stockmarket trading systems using artificial neural networks,” *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 6668–6680, 2009, doi:<https://doi.org/10.1016/j.eswa.2008.08.019>.
- [25] Lobato, I., “Series de tiempo univariantes: Modelos arma.”, <https://1library.co/document/myjv15my-series-de-tiempo-univariantes-modelos-arma.html> (visitado el 2022-11-30).
- [26] Becerra, N., “Unidad 1: Introducción al procesamiento de voz y técnicas asociadas,” 2021.

- [27] “Modelos de algoritmos genéticos y redes neuronales en la predicción de Índices bursátiles asiáticos,” Cuadernos de Economía, vol. 43, pp. 251 – 284, 2006.
- [28] Schnader, M. H., . S. H. O., “Evaluating predictions of change. the journal of business,” Expert Systems with Applications, vol. 63, p. 99–107, 1990, <http://www.jstor.org/stable/2353240>.
- [29] JJ, “Mae and rmse — which metric is better?,” <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d> (visitado el 2016-03-23).
- [30] N.N., “Mean absolute percentage error (mape).”, <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/> (visitado el 2022-11-28).
- [31] Chugh, A., “Mae, mse, rmse, coefficient of determination, adjusted r squared — which metric is better?,” <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e> (visitado el 2020-12-08).
- [32] Arias, E. R., “Ratio de calmar,” 2022, <https://economipedia.com/definiciones/ratio-de-calmar.html>.
- [33] Peña Vargas, T., “Modelando la dirección del precio y retornos de las acciones chilenas usando redes neuronales recurrentes - lstm,” pp. 29–31, 2022.
- [34] Seitz, S., “Probabilistic cusum for change point detection.”, <https://www.sarem-seitz.com/probabilistic-cusum-for-change-point-detection/> (visitado el 2022-08-04).
- [35] 2022, <https://www.financialpipeline.com/expert/bond-spreads/>.
- [36] , <https://learn.robinhood.com/articles/5fFx966IXNQoCCZLqCTxx/what-is-a-market-index/>.
- [37] Rivero, R., “¿qué es la capitalización de mercado? - definición y ejemplos,” 2023, <https://admiralmarkets.com/es/education/articles/shares/capitalizacion-mercado>.
- [38] , <https://www.estrategiasdeinversion.com/herramientas/diccionario/analisis-tecnico/volumen-de-negociacion-en-bolsa-t-394>.
- [39] Habiger, S., “Book value per share: Definition, formula amp; example,” 2023,
- [40] De Mendon, F. J., “El exponente de hurst: ¿conocías esta técnica estadística?: Investing.com,” 2023, <https://es.investing.com/analysis/el-exponente-de-hurst-conocias-esta-tecnica-estadistica-200458826>.
- [41] Schaap, C., “Adx: The trend strength indicator,” 2022, <https://www.investopedia.com/articles/trading/07/adx-trend-indicator.asp#citation-4>.
- [42] Kuepper, J., “Cboe volatility index (vix): What does it measure in investing?,” 2023, <https://www.investopedia.com/terms/v/vix.asp>.
- [43] Team, Z. M. R., “Stock analyst: What is it? and how to become one?,”
- [44] Murphy, C., “Put-call ratio meaning and what it says about how to gauge the market,” 2023, <https://www.investopedia.com/ask/answers/06/putcallratio.asp>.
- [45] Fernando, J., “Bid and ask definition, how prices are determined, and example,” 2023, <https://www.investopedia.com/terms/b/bid-and-ask.asp>.
- [46] Rodrigo, J. A.,

- [47] Gutierrez, J. M., “¿qué es la autocorrelación? test de durbin-watson,” 2022, <https://toeconometria.com/autocorrelacion-dw/>.
- [48] Aleksander Dietrichson, P., “Métodos cuantitativos.”, <https://bookdown.org/dietrichson/metodos-cuantitativos/test-de-normalidad.html>.
- [49] 2021, <https://statologos.com/prueba-pagana-breusch/>.
- [50] Bevans, R., “Akaike information criterion: When amp; how to use it (example),” 2023, <https://www.scribbr.com/statistics/akaike-information-criterion/>.
- [51] Moradi, A., “Xibalbas/signal_backtester: Tiny backtester to backtest generated signals,” 2022, https://github.com/xibalbas/signal_backtester/tree/master.
- [52] Santos, M. S. y Iraola, M. A., “.international encyclopedia of the social sciences. . encyclopedia.com. 22 aug. 2023 lt;https://www.encyclopedia.comgt;.”, 2023, <https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/market-fundamentals>.
- [53] Bohl, L. y Frederick, R., “Cómo elegir acciones usando el análisis fundamental y técnico.”, <https://international.schwab.com/es/content/how-to-pick-stocks-using-fundamental-and-technical-analysis>.
- [54] FXCM Team, F., <https://www.fxcm.com/markets/insights/sentiment-analysis-what-is-it-why-do-you-need-it/>.
- [55] Canut, F., <https://www.forex.com/es-latam/news-and-analysis/volatilidad-en-mercados-financieros/>.
- [56] Communications, B., “Qué es la liquidez financiera y por qué es tan importante,” 2023, <https://www.bbva.com/es/salud-financiera/que-es-la-liquidez-financiera-y-por-que-es-tan-importante/>.
- [57] Hayes, A., “Behavioral finance: Biases, emotions and financial behavior.”, <https://www.investopedia.com/terms/b/behavioralfinance.asp>.

Anexo A

Simulación de Trading para MAE

A.1. Optimizando MAE

- SET 1

Tabla A.1: Top 10 acciones para el set 1

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
WMB	0.526	0.089	0.007	two_side_sl_trailing_reversed	1.134	3.247	515.812	-21.336
MRO	0.502	0.099	0.006	two_side_sl_tp_reversed	0.840	2.196	337.935	-53.341
STX	0.518	0.076	0.004	two_side_sl_trailing_reversed	1.101	4.721	351.460	-14.809
OKE	0.505	0.079	0.006	two_side_sl_trailing_reversed	0.866	1.190	174.251	15.914
WDC	0.507	0.076	0.003	two_side_sl_trailing_reversed	0.813	1.362	151.137	-5.140
BEN	0.464	0.100	0.006	two_side_sl_tp_reversed	0.996	2.391	125.356	-18.780
CF	0.495	0.104	0.008	one_side_sell_sl_trailing	0.581	0.595	49.810	-48.092
APA	0.483	0.086	0.007	one_side_sell_sl_trailing	0.585	0.685	52.074	-36.522
ALB	0.504	0.126	0.017	two_side_sl_tp_reversed	0.847	0.855	121.049	32.758
BKNG	0.507	0.101	0.010	two_side_sl_trailing_reversed	1.122	2.466	189.928	101.990

- SET 2:

Tabla A.2: Top 10 acciones para el set 2

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
PARA	0.519	0.081	0.007	two_side_sl_tp_reversed	0.685	1.617	403.456	-47.446
PVH	0.512	0.079	0.008	two_side_sl_trailing_reversed	0.872	2.031	382.689	-37.260
TFC	0.482	0.143	0.029	two_side_sl_tp_reversed	0.720	1.427	284.007	-5.112
CCL	0.487	0.071	0.005	two_side_sl_tp_reversed	0.346	0.665	181.500	-80.897
WYNN	0.483	0.094	0.014	two_side_sl_trailing_reversed	0.666	1.387	197.071	-41.973
UAL	0.506	0.085	0.010	two_side_sl_tp_reversed	0.356	0.655	152.916	-51.730
BA	0.461	0.082	0.010	two_side_sl_tp_reversed	0.474	0.864	150.553	-52.992
IRM	0.507	0.131	0.015	two_side_sl_trailing_reversed	1.154	2.416	239.679	55.010
EXPE	0.511	0.116	0.018	two_side_sl_tp_reversed	0.506	1.004	174.038	-7.739
RF	0.484	0.111	0.014	two_side_sl_tp_reversed	0.540	0.891	148.708	9.082

- SET 3:

Tabla A.3: Top 10 acciones para el set 3

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
PARA	0.500	0.088	0.008	two_side_sl_tp_reversed	0.519	0.567	544.562	-40.444
PVH	0.498	0.090	0.009	two_side_sl_trailing_reversed	0.565	0.736	498.027	-44.746
IRM	0.515	0.109	0.015	two_side_sl_trailing_reversed	0.783	0.714	541.413	75.863
VTR	0.496	0.100	0.015	two_side_sl_tp_reversed	0.504	0.385	315.942	-14.599
HUM	0.477	0.101	0.009	one_side_buy_sl_trailing	1.109	2.025	789.319	545.439
PENN	0.520	0.109	0.009	one_side_buy_sl_tp	0.503	0.438	440.264	229.845
WAB	0.516	0.094	0.008	two_side_sl_trailing_reversed	0.575	0.640	297.809	100.865
VFC	0.501	0.093	0.008	two_side_sl_trailing_reversed	0.496	0.599	218.780	31.153
APA	0.489	0.084	0.007	one_side_sell_sl_trailing	0.262	0.155	92.934	-61.531
RL	0.504	0.098	0.013	two_side_sl_trailing_reversed	0.288	0.190	112.637	-40.082

Tabla A.4: Estrategias en acciones que presentan mayor retorno sobre buy and hold

Estrategias	SET 1	SET 2	SET 3
one side buy stoploss take profit	6	24	17
one side buy stoploss trailing	17	47	27
one side sell stoploss take profit	1	1	1
one side sell stoploss trailing	10	5	5
two side stoploss take profit reversed	16	27	3
two side stoploss trailing reversed	18	13	8
Total	68	117	61

Anexo B

Simulación de Trading para DIS

B.1. Optimizando DIS

- SET 1

Tabla B.1: Top 10 acciones para el set 1

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
WMB	0.526	0.089	0.007	two_side_sl_trailing_reversed	1.134	3.247	515.812	-21.336
MRO	0.502	0.099	0.006	two_side_sl_tp_reversed	0.840	2.196	337.935	-53.341
STX	0.518	0.076	0.004	two_side_sl_trailing_reversed	1.101	4.721	351.460	-14.809
WDC	0.507	0.076	0.003	two_side_sl_trailing_reversed	0.813	1.362	151.137	-5.140
OKE	0.506	0.079	0.006	two_side_sl_trailing_reversed	0.844	1.019	165.160	15.914
BEN	0.464	0.100	0.006	two_side_sl_tp_reversed	0.996	2.391	125.356	-18.780
CF	0.495	0.104	0.008	one_side_sell_sl_trailing	0.581	0.595	49.810	-48.092
APA	0.483	0.086	0.007	one_side_sell_sl_trailing	0.585	0.685	52.074	-36.522
ALB	0.504	0.126	0.017	two_side_sl_tp_reversed	0.847	0.855	121.049	32.758
BKNG	0.507	0.101	0.010	two_side_sl_trailing_reversed	1.122	2.466	189.928	101.990

- SET 2

Tabla B.2: Top 10 acciones para el set 2

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
PVH	0.512	0.079	0.008	two_side_sl_trailing_reversed	0.872	2.031	382.689	-37.260
PARA	0.518	0.082	0.007	two_side_sl_trailing_reversed	0.810	2.082	295.924	-47.446
IRM	0.504	0.131	0.015	two_side_sl_trailing_reversed	1.311	3.167	328.492	55.010
CCL	0.487	0.071	0.005	two_side_sl_tp_reversed	0.346	0.665	181.500	-80.897
WYNN	0.483	0.094	0.014	two_side_sl_trailing_reversed	0.666	1.387	197.071	-41.973
UAL	0.506	0.085	0.010	two_side_sl_tp_reversed	0.356	0.655	152.916	-51.730
BA	0.461	0.082	0.010	two_side_sl_tp_reversed	0.474	0.864	150.553	-52.992
EXPE	0.511	0.116	0.018	two_side_sl_tp_reversed	0.506	1.004	174.038	-7.739
VTR	0.491	0.092	0.016	two_side_sl_tp_reversed	0.492	0.633	121.065	-17.382
TFC	0.471	0.145	0.025	two_side_sl_tp_reversed	0.537	0.975	146.069	9.077

- SET 3

Tabla B.3: Top 10 acciones del set 3

stock	DA	MAE	DIS	estrategias	Sharp ratio	Calmar ratio	retorno	Buy hold
PVH	0.498	0.090	0.009	two_side_sl_trailing_reversed	0.565	0.736	498.027	-44.746
IRM	0.515	0.109	0.015	two_side_sl_trailing_reversed	0.783	0.714	541.413	75.863
PARA	0.498	0.089	0.008	two_side_sl_trailing_reversed	0.529	0.630	357.436	-40.444
VTR	0.492	0.102	0.014	two_side_sl_tp_reversed	0.466	0.334	279.823	-12.726
HUM	0.477	0.101	0.009	one_side_buy_sl_trailing	1.109	2.025	789.319	545.439
PENN	0.520	0.109	0.009	one_side_buy_sl_tp	0.503	0.438	440.264	229.845
WAB	0.516	0.094	0.008	two_side_sl_trailing_reversed	0.575	0.640	297.809	100.865
VFC	0.501	0.093	0.008	two_side_sl_trailing_reversed	0.496	0.599	218.780	31.153
APA	0.489	0.084	0.007	one_side_sell_sl_trailing	0.262	0.155	92.934	-61.531
RL	0.504	0.098	0.013	two_side_sl_trailing_reversed	0.288	0.190	112.637	-40.082

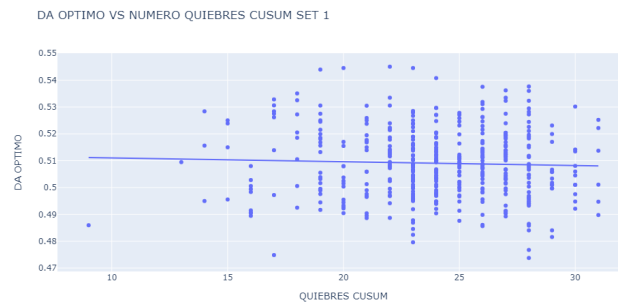
Tabla B.4: Estrategias en acciones que presentan mayor retorno sobre buy and hold

Estrategias	SET 1	SET 2	SET 3
one side buy stoploss take profit	4	23	16
one side buy stoploss trailing	17	47	28
one side sell stoploss takeprofit	1	1	1
one side sell stoploss trailing	10	5	5
two side stoploss take profit reversed	15	23	2
two side stoploss trailing reversed	18	18	9
Total general	65	117	61

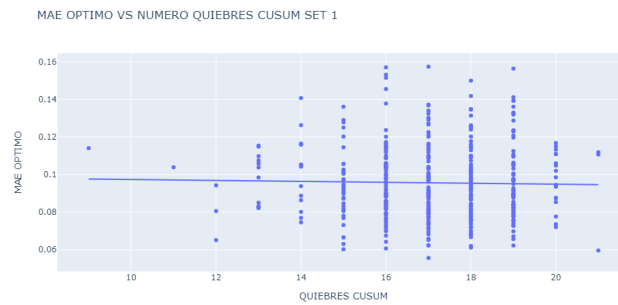
Anexo C

Cusum

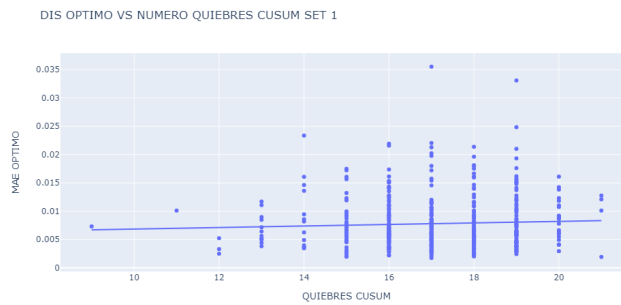
Para el set 1:



(a) Para DA



(b) Para MAE

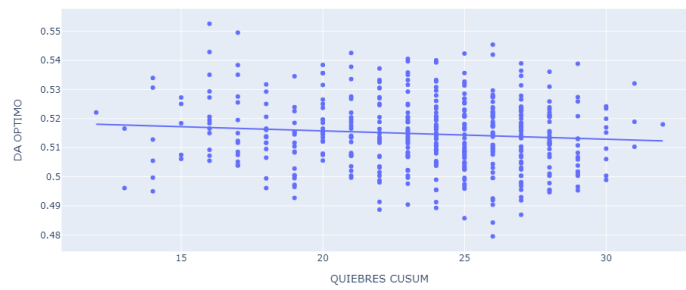


(c) Para DIS 1

Figura C.1: Cantidad de quiebres contra las distintas métricas.

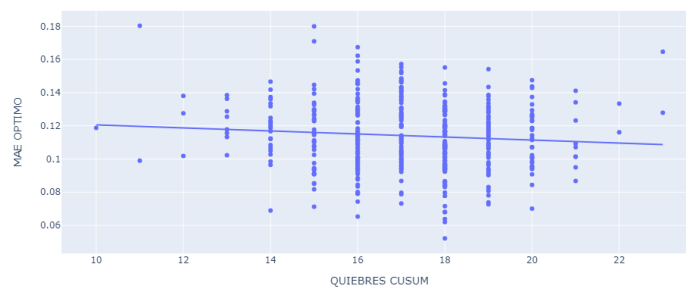
Para el set 2

DA OPTIMO VS NUMERO QUIEBRES CUSUM SET 2



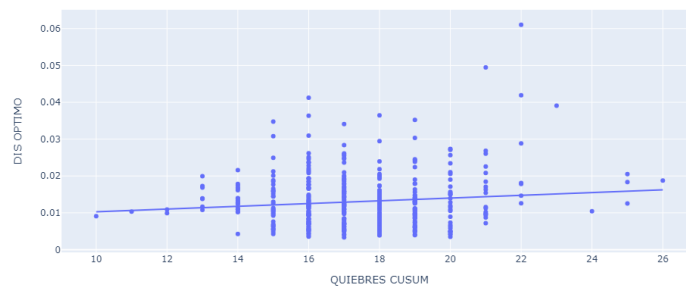
(a) Para DA

MAE OPTIMO VS NUMERO QUIEBRES CUSUM SET 2



(b) Para MAE

DIS OPTIMO VS NUMERO QUIEBRES CUSUM SET 2



(c) Para DIS 1

Figura C.2: Cantidad de quiebres contra las distintas métricas.

Anexo D

Simulación de trading

D.1. Regresiones para el set 1

Tabla D.1: Set 1 Determinantes para el DA

TABLE	REG_1	REG_2	REG_3	REG_4	REG_5	REG_6	REG_7	REG_8	REG_9	REG_10	REG_11	REG_12	REG_13	REG_14	REG_15
DEPVAR	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA
OBSERVATIONS	420	420	420	420	420	420	420	420	420	420	420	420	420	420	420
R2	57%	62%	55%	59%	7%	27%	4%	25%	56%	61%	56%	60%	56%	61%	62%
ADJUSTED R2	56%	58%	54%	56%	6%	21%	2%	19%	55%	57%	55%	57%	55%	57%	61.10%
CONST.	16.57 ***	17.62 ***	15.33 ***	16.67 ***	51.16 ***	48.95 ***	51.73 ***	49.18 ***	16.33 ***	17.32 ***	16.7 ***	17.97 ***	16.61 ***	17.49 ***	14.54 ***
ADX30D_TREND_DAYS.	-0.0007	-0.0006	-0.0006	-0.0003	-0.00009038	0.0004	0.0003	0.0004						-0.0008*	-0.0006
	0.001	0.001	0	0.001	0.001	0.001	0.001	0.001						0	0.001
ANALYST_DISPERSION_RMEAN.	0.004*	0.0052**			0.012***	0.0112***			0.0052**	0.0062***	0.0043*	0.0048**	0.0041*	0.0051**	0.004*
	0.002	0.002			0.003	0.003			0.002	0.002	0.002	0.002	0.002	0.002	0.002
ANALYST_NUMBER_RMEAN.	0.0008	0.0049	-0.0001	0.0049			-0.0084	-0.0021	0.00002768	0.0054	0.0017	0.007	-0.0012	0.0062	0.0062
	0.007	0.008	0.007	0.007			0.009	0.01	0.007	0.008	0.006	0.007	0.005	0.007	0.007
BIDASK_RMEAN.	2.4577	2.6019			-0.9859	-0.3406	-0.5409	-0.6058	2.2893	2.736			2.45	2.9084*	1.423
	1.627	1.674			2.331	2.249	2.411	2.296	1.633	1.664			1.628	1.668	1.513
HURST_RANDWALK_DAYS.	0.0004*	0.0004											0.0004*	0.0004	0.0004*
	0	0											0	0	0
MARKETCAP_RMEAN.	2.363E-07	-6.193E-08	-3.589E-08	-0.000000126					1.307E-07	-0.000000135					
	0.00000102	0.00000105	0.00000102	0.00000105					0.000000984	0.000001					
MARKET_BETA.	2.9627**	2.0007**					4.6165**	5.4643*	-0.271*	-0.4293*					-0.375***
	1.35	1.598					1.957	2.161	0.159	0.222					0.143
PBPERSHARE_RMEAN.	0.0019	0.0026**			0.0036**	0.0043***	0.0042**	0.0039**	0.0019	0.0027**	0.0025**	0.003**	0.002*	0.0027**	0.002*
	0.001	0.001			0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001
PRICE_CUSUM_BREAKS.	-0.0014	-0.0056	-0.0056	-0.0087	-0.0467	-0.0314	-0.0435	-0.0312	-0.0109	-0.0104			-0.0124	-0.0115	
	0.022	0.022	0.022	0.023	0.032	0.03	0.033	0.031	0.022	0.022			0.022	0.022	
PUTCALLRATIO_RMEAN.	0.0047	-0.0091					0.0029	0.0068	-0.0062	-0.0205					0.017
	0.036	0.037					0.052	0.049	0.036	0.036					0.03
R_MSTD_LT.	-1.6556	-1.2787									-2.081**	-1.7044			
	1.097	1.153									1.021	1.092			
R_MSTD_ST.	1.8811	1.4194	-0.1519	-0.1691	0.4224***	0.231					2.0387*	1.6242			
	1.204	1.273	0.122	0.142	0.155	0.178					1.086	1.173			
SPREAD_BETA.	2.5354**	2.3624**	2.9321***	2.6637***			-0.6947	-0.4708	2.3031**	2.2411**			2.2116**	2.1196**	2.308**
	0.987	1.019	0.956	0.992			1.423	1.365	0.969	0.994			0.947	0.985	0.988
STREAK_HMEAN.	17.2266***	16.4036***	17.9769***	16.9282***					17.498***	16.649***	16.9799***	15.9113***	17.198***	16.2662***	17.843***
	0.85	0.924	0.837	0.911					0.815	0.904	0.829	0.907	0.814	0.905	0.773
VIX_BETA.	35.74**	26.5247					44.3501**	52.4548**							
	14.211	16.555					20.188	22.259							
VOLUME_CV.	0.001	0.001			0.0027	0.0019	0.0024	0.0026			0.0016	0.0017			
	0.002	0.002			0.002	0.002	0.002	0.002			0.002	0.002			
W_SIZE															0*
															0
P_LAGS															0.032***
															0.004
INDUSTRIES	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

D.2. Regresiones para el set 2

Tabla D.2: Set 2 determinantes para el DA

TABLE	REG_1	REG_2	REG_3	REG_4	REG_5	REG_6	REG_7	REG_8	REG_9	REG_10	REG_11	REG_12	REG_13	REG_14	REG_15
DEPVAR	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA
OBSERVATIONS	424	424	424	424	424	424	424	424	424	424	424	424	424	424	424
R2	0.515	0.576	0.506	0.565	0.049	0.203	0.11	0.223	0.504	0.563	0.509	0.571	0.496	0.561	0.603
ADJUSTED R2	0.496	0.532	0.498	0.531	0.033	0.14	0.088	0.155	0.492	0.525	0.5	0.536	0.487	0.527	0.594
CONST	16.89 ***	16.49 ***	17.07 ***	16.93 ***	53.01 ***	49.78 ***	52.63 ***	49.82 ***	17.74 ***	17.26 ***	18.4805	18.48 ***	16.84 ***	16.96 ***	14.3288 ***
ADX30D_TREND_DAYS.	2.072	2.024	2.02	1.985	0.907	0.895	0.921	0.929	2.056	2.008	1.832	1.802	2.023	1.953	1.891
	-0.0007	-0.0007	-0.0008	-0.0009	-0.0002	-0.00003499	-0.0001	-0.0002							
	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001							
ANALYST_DISPERSION_RMEAN.	-0.0018	0.0011			0.0033	0.0032			-0.0019	-0.0000524			-0.0026	0.0001	
	0.003	0.003			0.004	0.004			0.003	0.003			0.003	0.003	
ANALYST_NUMBER_RMEAN.	-0.0044	0.0004	-0.0122	-0.0056			-0.0106	-0.004	-0.0092	-0.0007	-0.0031	0.0039			-0.009
	0.009	0.01	0.007	0.009			0.011	0.012	0.008	0.009	0.007	0.008			0.007
BIDASK_RMEAN.	0.0647	-0.0416			0.0434	-0.0882	-0.0269	-0.4059	-0.0391	-0.2382			0.2526	0.0291	0.1868
	0.536	0.539			0.722	0.715	0.718	0.718	0.531	0.537			0.521	0.528	0.476
HURST_RANDWALK_DAYS.	-0.0000371	-0.0003											-0.0001	-0.0004	
	0	0											0	0	
MARKETCAP_RMEAN.	3.965E-07	4.049E-07	2.586E-07	2.929E-07					4.299E-07	3.564E-07					
	0.000000452	0.000000454	0.00000044	0.000000439					0.000000431	0.000000436					
MARKET_BETA.	-0.2575	1.1573					0.5654	0.4726	-0.5056 ***	-0.1879	-0.3097	1.1892			-0.3792 **
	0.687	0.822					0.891	1.079	0.186	0.249	0.671	0.809			0.167
PBPERSHARE_RMEAN.	-0.0003	-0.0003			0.0012	0.0017	0.0014	0.0015	-0.00003884	0.00003377			-0.00007581	0.00001058	
	0.001	0.001			0.002	0.002	0.002	0.002	0.001	0.001			0.001	0.001	
PRICE_CUSUM_BREAKS.	0.0368 *	0.0282	0.0323	0.0224	-0.0434	-0.0245	-0.0306	-0.03	0.027	0.0169			0.0368 *	0.0259	0.0357 *
	0.022	0.022	0.022	0.021	0.03	0.029	0.029	0.029	0.021	0.021			0.022	0.021	0.019
PUTCALLRATIO_RMEAN.	0.0664	0.0655					0.1554 **	0.165 **	0.071	0.0634	0.0704	0.0656			0.044
	0.051	0.052					0.068	0.069	0.051	0.052			0.05	0.051	0.045
R_MSTD_LT.	1.9177 **	1.6458 *											1.8124 **	1.636 *	
	0.888	0.889											0.877	0.874	
R_MSTD_ST.	-2.1663 **	-1.7402 *	-0.3054 **	-0.1448	-0.554 ***	-0.1152							-2.1569 **	-1.8438 *	
	0.948	0.96	0.127	0.152	0.153	0.19							0.937	0.944	
SPREAD_BETA.	-1.6037 *	-1.9186 **	-1.4853 **	-1.7077 *			-1.3366	-1.9363	-1.0831	-1.5148	-1.5651 **	-1.8217 **			-1.3063 *
	0.798	0.947	0.752	0.908			1.029	1.249	0.757	0.924	0.757	0.908			0.667
STREAK_HMEAN.	16.9461 ***	16.7768 ***	17.1252 ***	16.7638 ***					16.6918 ***	16.458 ***	16.7204 ***	16.6779 ***	17.0427 ***	16.7473 ***	17.817 ***
	0.937	0.952	0.899	0.937					0.919	0.95	0.914	0.933	0.911	0.934	0.842
VIX_BETA.	0.2891	14.4307					14.3878 *	5.3436			-1.0444	13.5734	6.7288 ***	4.9116	
	6.988	8.959					8.691	11.109			6.768	8.778	1.647	2.401	
VOLUME_CV.	0.0015	0.0005			0.005 **	0.0031	0.0019	0.0009							
	0.002	0.002			0.002	0.002	0.002	0.002							
W_SIZE															0.0014 ***
															0
P_LAGS															0.0365 ***
															0.004
INDUSTRIES	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	no

D.3. Regresiones para el set 3

Tabla D.3: Set 3 Determinantes para el DA

TABLE	REG_1	REG_2	REG_3	REG_4	REG_5	REG_6	REG_7	REG_8	REG_9	REG_10	REG_11	REG_12	REG_13	REG_14	REG_15	
DEPVAR	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	DA	
OBSERVATIONS	424	424	424	424	424	424	424	424	424	424	424	424	424	424	424	
R2	0.551	0.583	0.536	0.571	0.026	0.201	0.042	0.199	0.546	0.58	0.537	0.571	0.54	0.575	0.617	
ADJUSTED R2	0.533	0.54	0.528	0.537	0.009	0.137	0.019	0.128	0.535	0.543	0.528	0.536	0.531	0.54	0.611	
CONST.[ICOEFF]	15.1951	15.5534	15.6969	16.3526	51.6035	48.7706	52.1769	49.1768	15.9796	16.205	16.2293	16.7284	16.152	16.3859	12.667***	
ADX30D_TREND_DAYS.	-0.0001	-0.0001	-1.9E-05	-2.9E-05	0.0001	0.0001	0.0002	0.0001								
	0	0	0	0	0	0	0	0								
ANALYST_DISPERSION_RMEAN.	-0.0006	0.0011			0.0063*	0.0061*				-0.0007	0.0009	-0.0006	0.0007	-0.0015	0.0002	
	0.003	0.003			0.004	0.004				0.002	0.003	0.003	0.003	0.002	0.003	
ANALYST_NUMBER_RMEAN.	0.0026	0.0056	-8.7E-05	0.0025			-0.0147*	-0.006	0.0002	0.0039	0.0027	0.0058	-0.0034	0.0035		
	0.006	0.007	0.005	0.006			0.008	0.009	0.006	0.007	0.006	0.006	0.004	0.005		
BIDASK_RMEAN.	1.6905***	1.6788***			0.5971	0.697	0.6732	0.5436	1.6287***	1.632***			1.6476***	1.7215***	1.7263***	
	0.563	0.589			0.812	0.797	0.811	0.802	0.551	0.579			0.552	0.579	0.503	
HURST_RANDWALK_DAYS.	6.94E-05	3.93E-06											9.16E-06	-4.7E-05		
	0	0											0	0		
MARKETCAP_RMEAN.	-6.2E-08	4.96E-08	-1.8E-07	-6.8E-08						1.85E-08	1.7E-07					
	4.63E-07	4.8E-07	4.58E-07	4.7E-07						4.41E-07	4.54E-07					
MARKET_BETA.	0.8139	0.9086					0.6849	0.0166	-0.2823**	-0.3562*	0.5636	0.8383				
	0.75	0.927					1.004	1.199	0.126	0.185	0.666	0.858				
PBPERSHARE_RMEAN.	-0.0003	-0.0002			0.0009	0.0011	0.0012	0.001	-0.0003	-0.0002			-0.0005	-0.0003		
	0.001	0.001			0.002	0.002	0.002	0.002	0.001	0.001			0.001	0.001		
PRICE_CUSUM_BREAKS.	0.0034	0.003	0.0028	0.0015	-0.0266*	-0.0132	-0.0267*	-0.0171	-7.7E-05	0.0005	0.0023	-0.0003	-0.0039	-0.0011		
	0.011	0.011	0.011	0.011	0.016	0.015	0.016	0.015	0.011	0.011	0.011	0.011	0.011	0.011	0.011	
PUTCALLRATIO_RMEAN.	0.0198	0.0181					0.0198	0.0474	0.0251	0.0236	0.0304	0.03			0.0113	
	0.035	0.036					0.05	0.049	0.034	0.035	0.034	0.036			0.026	
R_MSTD_LT.	-1.0895	-0.6727													-0.1673**	
	1.034	1.08													0.071	
R_MSTD_ST.	1.0608	0.6034	-0.2156**	-0.2534**	0.0109	0.1031										
	1.107	1.161	0.091	0.114	0.121	0.149										
SPREAD_BETA.	1.493	2.2135	1.3298	2.0908			-1.0492	-0.6433	1.3402	2.2856			0.7681	1.6036	1.7066	
	1.399	1.548	1.383	1.523			1.982	2.099	1.373	1.529			1.347	1.494	1.238	
STREAK_HMEAN.	17.633***	17.1859***	17.5368***	16.9144***						17.3928***	16.9174***	17.1467***	16.6445***	17.336***	16.7468***	18.8536***
	0.861	0.926	0.832	0.912						0.817	0.904	0.817	0.906	0.826	0.903	0.749
VIX_BETA.	10.1565	11.6964					8.5853	-1.1162			8.3434	12.2305				
	7.733	9.832					9.914	12.169			6.586	8.77				
VOLUME_CV.	0.0007	0.0009			0.0032*	0.0025	0.0011	0.0015			0.0007	0.0007				
	0.001	0.001			0.002	0.002	0.002	0.002			0.001	0.001				
W_SIZE																
P_LAGS															0.0258***	
															0.003	
INDUSTRIES	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	