



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Conformación Automática de Portafolios de Inversión usando Analítica Financiera

Laura Cristina Echeverri Sánchez

Universidad Nacional de Colombia
Facultad de Minas, Área Curricular de Sistemas e Informática
Medellín, Colombia

2021

Conformación Automática de Portafolios de Inversión usando Analítica Financiera

Laura Cristina Echeverri Sánchez

Tesis de investigación presentada como requisito parcial para optar al título de:
Magister en Ingeniería – Ingeniería de Sistemas

Director:

Juan David Velásquez Henao, PhD

Línea de Investigación:

Analítica Predictiva

Grupo de Investigación:

Big Data & Data Analytics

Universidad Nacional de Colombia

Facultad de Minas, Área Curricular de Sistemas e Informática

Medellín, Colombia

2021

(Dedicatoria o lema)

Si buscas resultados distintos, no hagas siempre lo mismo.

Albert Einstein

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Laura Cristina Echeverri Sánchez

27/09/2021

Agradecimientos

A la Universidad Nacional de Colombia por brindarme la oportunidad de realizar mi maestría a través de una beca. ¡El amor y gratitud hacia mi alma mater es inmensurable!

A Juan David Velásquez Henao, PhD, director de esta tesis, por su guía, sus enseñanzas, su apoyo en todo momento y por transmitirme seguridad. Su voto de confianza en mí, hizo posible la realización de este trabajo.

A mi familia, mis padres y hermano, por su amor y apoyo incondicional.

A todo aquel que en algún momento me brindó un consejo y su buena energía para culminar esta maestría.

Resumen

Conformación Automática de Portafolios de Inversión usando Analítica Financiera.

En este trabajo se presenta un prototipo de simulación para evaluar diferentes estrategias de Comercio Algorítmico en el mercado financiero colombiano; esto con el fin de analizar si es posible incorporar este tipo de estrategias por parte de los inversionistas. Para construir las estrategias, se hacen uso de diversos tipos de modelos de Inteligencia Artificial, como por ejemplo redes neuronales, bosques aleatorios y regresión logística, los cuales predicen la tendencia del precio del día siguiente. Estas predicciones son transformadas en señales de compra y venta de las acciones que permiten la conformación diaria del portafolio. Las diferentes estrategias varían en cuanto al tipo de modelo entrenado para cada activo, el subconjunto de acciones seleccionado y otros parámetros que se dan en la negociación y que dependen exclusivamente de la aversión al riesgo del inversionista, tal como el porcentaje invertido en cada movimiento y la pérdida máxima aceptada. Las diferentes simulaciones permiten establecer la estrategia que logra la mayor rentabilidad para el inversionista, que en el escenario planteado en este trabajo consta de la selección de 11 acciones y un tipo de modelo diferente para cada activo según su mejor desempeño predictivo. Dicha estrategia alcanza una rentabilidad de 78% sobre la inversión. Los resultados de esta estrategia automática de negociación fueron comparados con la rentabilidad generada por la estrategia tradicional de conformación de portafolio Markowitz, la cual genera un 5% de pérdida. Al contrastar estos resultados se aprecian las bondades que trae para el inversionista implementar una estrategia automática de negociación basada en la predicción de la dirección del precio de las acciones.

Palabras clave: comercio algorítmico, inteligencia artificial, aprendizaje supervisado, mercado financiero.

Abstract

Automatic Conformation of Investment Portfolios using Financial Analytics.

In this work a simulation prototype is presented to evaluate different Algorithmic Trading strategies in the Colombian financial market; the purpose is to analyze the possibility to incorporate this type of strategy by investors. In order to build the strategies, various types of Artificial Intelligence models are applied, such as neural networks, random forests and logistic regression, which predict the price trend of the next day. These predictions are transformed into buy and sell signals for the stocks that allow the daily formation of the portfolio. The different strategies vary in terms of the type of model trained for each asset, the selected subset of stocks and other parameters that occur in the negotiation and that depend exclusively on the investor's aversion to risk, such as the percentage invested in each movement and the maximum accepted loss. The different simulations make it possible to establish the strategy that achieves the highest profitability for the investor, which in the scenario proposed in this work consists of the selection of 11 stocks and a different type of model for each asset according to its best predictive performance. This strategy achieves a 78% return on investment. The results of this automatic trading strategy were compared with the profitability generated by the traditional Markowitz portfolio formation strategy, which generates a 5% loss. When comparing these results, the benefits that the investor brings to implement an automatic negotiation strategy based on the prediction of the direction of the share price can be appreciated.

Keywords: algorithmic trading, artificial intelligence, supervised learning, financial market.

Contenido

	Pág.
Resumen	IX
Lista de figuras.....	XIII
Lista de tablas	XIV
1. Introducción	15
1.1 Definición del problema de investigación.....	18
1.2 Revisión de Literatura	20
1.2.1 Metodología.....	21
1.2.1.1 Proceso de búsqueda.....	21
1.2.1.2 Preguntas a resolver con el MSL.....	22
1.2.2 Resultados obtenidos	23
1.2.3 Respuestas a las preguntas de investigación	23
1.3 Discusión	37
1.4 Hipótesis	40
1.5 Objetivos.....	41
1.5.1 Objetivo General.....	41
1.5.2 Objetivos Específicos.....	41
1.6 Mapa del documento.....	41
2. Aproximación Metodológica Propuesta	42
2.1 Estrategia propuesta para predecir el movimiento del precio de las acciones ..	42
2.1.1 Selección de acciones	42
2.1.2 Selección de periodo de observación	43
2.1.3 Definición del problema de pronóstico	44
2.1.4 Definición de variables.....	45
2.1.4.1 Variables independientes.....	45
2.1.4.2 Variable dependiente (respuesta)	46
2.1.5 Técnicas de pronóstico	47
2.1.6 Evaluación de los modelos	48
2.2 Estrategia propuesta para la construcción del portafolio.....	49
2.2.1 Utilización del modelo predictivo de la dirección del precio para tomar decisiones de compra y venta de acciones	49
2.2.2 Priorización de compra de acciones	51
2.2.3 Definición de parámetros en la negociación.....	52
2.2.3.1 Precio de compra y venta de acciones	53
2.2.3.2 Costo de transacción	53
2.2.3.3 Capital inicial	53

2.2.3.4	Porcentaje del capital disponible a invertir en cada operación.....	53
2.2.3.5	Pérdida máxima	54
2.3	Conclusiones	54
3.	Resultados y Discusión.....	56
3.1	Definición del problema de negocio	56
3.2	Definición del problema de analítica.....	57
3.3	Información utilizada	58
3.4	Preparación de datos e ingeniería de características	59
3.4.1	Entendimiento de los datos	59
3.4.2	Preparación de los datos.....	63
3.4.2.1	Variables y retardo	63
3.4.2.2	Normalización de las variables predictoras.....	64
3.4.2.3	Definición de variable respuesta.....	64
3.5	Metaheurística propuesta para usar los modelos de pronóstico en la conformación del portafolio.....	66
3.6	Resultados obtenidos y discusión	69
3.6.1	Modelo de simulación.....	69
3.6.2	Selección de características y retardo	69
3.6.3	Estimación de los parámetros del modelo para todas las acciones del portafolio.....	73
3.6.3.1	Pronóstico usando el mismo tipo de modelo para todo el portafolio	73
3.6.3.2	Pronóstico usando diferente tipo de modelo para cada acción	76
3.6.3.3	Selección de un subconjunto de acciones para el portafolio.....	80
3.6.3.4	Evaluación del porcentaje de capital a invertir en cada operación.....	84
3.6.4	Comparación de la estrategia trading algorítmico vs. la estrategia mínima varianza	88
3.6.4.1	Supuestos para la conformación de portafolio según modelo Markowitz: 89	
3.6.4.2	Resultados obtenidos.....	90
3.7	Conclusiones	91
4.	Conclusiones	93
4.1	Respuesta a las preguntas de investigación	93
4.2	Cumplimiento de los objetivos.....	94
4.2.1	Objetivo general	94
4.2.2	Objetivo Específico 1	95
4.2.3	Objetivo Específico 2.....	97
4.2.4	Objetivo Específico 3.....	98
Bibliografía		100

Lista de Figuras

	Pág.
Figura 1-1: Publicaciones por año.	24
Figura 1-2: Principales autores de AT con su número de publicaciones por año.	25
Figura 1-3: Cantidad de citas por año para los 6 autores más citados.	26
Figura 1-4: Coautoría entre autores principales.	28
Figura 1-5: Publicaciones por año de las principales revistas.	29
Figura 1-6: Citaciones por año para las 10 fuentes más citadas.	30
Figura 1-7: Publicaciones por país.	32
Figura 1-8: Publicaciones por tipo de activo.	33
Figura 1-9: Cantidad de artículos publicados por año para las principales técnicas.	34
Figura 1-10: Cantidad de apariciones de palabras clave en artículos sobre AT antes de 2017 vs 2017-2021.	35
Figura 2-1: Periodo de análisis para entrenar y evaluar la predicción de la dirección del precio.	44
Figura 3-1: Histórico precio diario de las acciones.	59
Figura 3-2: Índice Colcap vs. ETF del índice S&P (SPY).	61
Figura 3-3: Composición de las clases 0 y 1 en cada acción en el periodo 2013-2019.	62
Figura 3-4: Modelo de simulación propuesto para la conformación del portafolio.	69
Figura 3-5: Comportamiento COLCAP vs ECOPETROL y PFBCOLOM.	71
Figura 3-6: Tipos de modelo con el número de acciones que logra el mejor desempeño predictivo.	76
Figura 3-7: Comportamiento valor portafolio en el tiempo.	79
Figura 3-8: Ganancia neta por acción.	80
Figura 3-9: Evolución portafolio con 11 acciones.	83
Figura 3-10: Ganancia por acción del portafolio conformado por 11 acciones.	84
Figura 3-11: Rentabilidad de los portafolios según porcentaje invertido en cada operación.	86
Figura 3-12: Número de acciones de los portafolios según porcentaje invertido en cada operación.	87
Figura 3-13: Ejemplo frontera eficiente.	89
Figura 3-14: Valor del portafolio con estrategia de trading automático vs. Markowitz.	91

Lista de Tablas

	Pág.
Tabla 1-1: Número total de documentos escritos por un mismo autor.....	27
Tabla 1-2: Artículos más citados.....	31
Tabla 1-3: Grupo de términos obtenidos de la matriz de concurrencia de palabras clave	37
Tabla 2-1: Acciones que han pertenecido al COLCAP en el periodo 2013-2020.	42
Tabla 2-2: Variables seleccionadas para predecir dirección del precio (Rankia, 2020, BVC, s.f.-a).	45
Tabla 2-3: Esquema representativo de las variables para pronosticar la dirección del precio en t+1.....	46
Tabla 2-4: Ejemplo de la definición de la variable respuesta.....	47
Tabla 2-5: Matriz de confusión.....	48
Tabla 2-6: Ejemplo de las decisiones de compra y venta de las acciones a partir de la predicción de la dirección del precio.	50
Tabla 2-7: Priorización de acciones para efectuar operaciones en el portafolio.	52
Tabla 3-1: Ejemplo construcción variable respuesta.....	65
Tabla 3-2: Resultados de combinaciones de variables y retardo para la predicción de la dirección del precio de ECOPETROL y PFBCOLOM.	72
Tabla 3-3: Ganancia al usar un mismo tipo de modelo en cada acción.....	74
Tabla 3-4: Precisión balanceada por tipo de modelo y por acción.	74
Tabla 3-5: Mejor tipo de modelo por acción para predecir la dirección del precio.	77
Tabla 3-6: Simulación portafolio utilizando diferentes modelos.....	78
Tabla 3-7: Priorización de activos según ganancia para incluir en el portafolio.....	81
Tabla 3-8: Rentabilidad del portafolio usando diferente número de acciones.....	82
Tabla 3-9: Simulación portafolio con diferente porcentaje a invertir en cada operación.	85
Tabla 3-10: Rentabilidad portafolio con estrategia de trading automático vs. Markowitz	90

1.Introducción

El acelerado desarrollo de la tecnología ha transformado el funcionamiento de los mercados haciendo que las transacciones sean mucho más rápidas y efectivas (Bahadur, 2015). Particularmente, la adopción de las nuevas tecnologías como las TIC, ha cambiado la dinámica de los mercados financieros y como tal ha contribuido a su expansión y desarrollo (Seo & Chai, 2013; Lechman & Marszk, 2017). En especial, el comercio algorítmico ha sido señalado como uno de los procesos más influyentes y revolucionarios (Nettles et al., 2015).

El comercio automatizado (AT) ha sido definido como “la actividad de comprar y vender instrumentos financieros con el propósito de obtener ganancias, a través del uso de reglas de negociación automatizadas” (Azzini & Tettamanzi, 2008, p. 62). En este sentido, en el AT se hace uso de algoritmos sofisticados, por lo que también se conoce con el término de “Comercio algorítmico”. A través de la implementación de estos algoritmos es posible tomar ciertas decisiones automáticamente en el ciclo comercial, incluido el análisis pre-trade (análisis de datos), la generación de señales de negociación (recomendaciones de compra y venta) y la ejecución comercial (gestión de pedidos) (Hu et al., 2015). Hogenboom (2012) ha manifestado que, debido a su alta productividad y a sus costos relativamente bajos, el comercio algorítmico se ha vuelto cada vez más popular y ha contribuido a generar un mayor nivel de integración y globalización de los mercados de capitales (Panourgias, 2015).

En el Siglo XX, el mercado financiero era caracterizado por tener exclusivamente interacción e intervención humana (Farjam & Kirchkamp, 2018). No obstante, en las últimas décadas este mercado ha evolucionado pasando de ser un mercado predominantemente manual a un mercado altamente automatizado y competitivo (Litzenberger, Castura & Gorelick, 2012), el cual es operado por máquinas de software automatizadas que comercian en latencias de milisegundos (Ruta, Ruta & Cen, 2017) y que operan con poca o ninguna supervisión humana (Mankad, Michailidis & Kirilenko, 2013). Para lograr lo

anterior, en muchos casos se implementan herramientas de inteligencia artificial (IA) (Lima & Sichman, 2007) y aprendizaje estadístico para la toma de decisiones a una frecuencia muy alta (Kablan, 2009). En este contexto, se ha señalado que “el aumento de la penetración de las TIC ha transformado la infraestructura de los mercados financieros” (Lechman & Marszk, 2017, p. 7), y gracias a esto, en el siglo XXI el mercado financiero se ha posicionado como un mercado híbrido, en el que intervienen tanto las computadoras como los seres humanos (Farjam & Kirchkamp, 2018).

No obstante, aunque en la literatura se ha manifestado que los mercados financieros modernos están dominados por herramientas algorítmicas de negociación (Raudys, Plikynas & Masteika, 2014), esto no sucede con paridad en todos los mercados del mundo. Particularmente, en algunas regiones como Estados Unidos, Europa y Asia-Pacífico, la mayor parte del volumen negociado se da a través de sistemas algorítmicos (Kong, 2013), pero, la situación en los países emergentes es diferente. Existe una brecha con respecto a los desarrollos que han hecho las grandes economías en temas de AT, por lo que llevar a cabo estrategias de negociación automatizadas representa un desafío actual para los inversores de países subdesarrollados (Khoury & Gorse, 2015).

Aquellos países que le han apostado a la implementación de AT en sus mercados financieros han hecho uso de diversas técnicas; entre las más populares se encuentra la Inteligencia Artificial (IA) (Lima & Sichman, 2007; Dejavite & Lima, 2010), la cual ha generado múltiples vías de investigación en el mundo de las finanzas y la economía (Maringer & Ramtohul, 2010), donde los profesionales del Trading la han utilizado para generar modelos de rendimiento ajustado al riesgo (Vella & Ng, 2014). Se ha expuesto que la IA ejerce gran influencia en los mercados financieros (Parkes & Wellman, 2015), por lo que el AT representa un área de importante aplicación para dicha técnica (Cartlidge, Szostek, De Luca & Cliff, 2012). Los estudios realizados hasta el momento han evidenciado la idoneidad de los métodos de IA en la aplicación de AT ya que dichas técnicas tienen alta capacidad para procesar relaciones no lineales, aprenden fácilmente y evolucionan con el tiempo y brindan información para la toma de decisiones (Hi'ovská & Koncz, 2012).

De esta manera, se han estudiado e implementado varias técnicas de IA tales como redes neuronales artificiales, sistemas expertos, aprendizaje automático, algoritmos genéticos,

lógica difusa, sistemas multiagente, sistemas basados en reglas, redes bayesianas y minería de texto, entre otros. A partir de esto, se han realizado una diversidad de investigaciones alrededor del comercio automatizado en donde se han abordado diferentes métodos para construir algoritmos de negociación, y como se mencionó anteriormente, esto se ha realizado especialmente en países desarrollados.

Con este contexto, en la investigación asociada a AT existe un vacío en cuanto al desarrollo de estudios enfocados en economías emergentes, por lo que resulta relevante realizar investigaciones en torno a la implementación de sistemas automáticos de trading para operar dentro de los mercados de capitales de países subdesarrollados, como por ejemplo para el caso colombiano. Esto resulta valioso para dichas economías, ya que las impulsaría a ser más competitivas en el mercado financiero al contar con estrategias de negociación algorítmicas que disminuyan el riesgo inherente a las operaciones netamente manuales.

En este capítulo se formula el proyecto de investigación abordado en esta tesis, en el cual se pretende construir un prototipo de simulación para evaluar estrategias de Comercio Algorítmico en el mercado financiero colombiano, lo que permitirá la conformación automática de portafolios de inversión. Esto se llevará a cabo mediante el uso de Analítica Financiera, particularmente con el desarrollo de modelos construidos con técnicas de Inteligencia Artificial que permitan pronosticar el movimiento o la dirección del precio de las acciones y de esta manera sirvan de insumo para tomar decisiones automáticas de compra y venta de activos, que aumenten la rentabilidad.

El resto de este capítulo está organizado como sigue: en la Sección 1.1 se presenta el problema de investigación a partir del cual se origina la razón de ser de este trabajo, seguidamente en la Sección 1.2 se realiza la revisión de literatura sobre trading algorítmico, lo cual desencadena la Sección 1.3, donde se plantea la discusión alrededor de los resultados arrojados por dicha revisión. Finalmente, en la Sección 1.3 y 1.4 se presenta la Hipótesis y los objetivos de este trabajo de investigación, donde se dará cumplimiento a estos planteamientos en la Sección 2 (metodología) y la Sección 3 (resultados).

1.1 Definición del problema de investigación

En las diferentes transacciones que se dan en el mercado financiero existe un riesgo inherente asociado a las estrategias tradicionales frente a la posibilidad de elegir incorrectamente el momento para abrir o cerrar posiciones. Esto ha generado la necesidad del desarrollo de nuevas alternativas de inversión para la gestión de dicho riesgo, por lo que muchos mercados a nivel global han implementado y desarrollado sistemas de negociación automatizados, que han sido considerados como estrategias neutrales del riesgo (Gloukhov, Ilin, Kuposov & Levina, 2014).

Para aquellos países que han implementado AT en sus mercados financieros, la intuición humana y el conocimiento complejo adquirido durante años para tomar decisiones comerciales óptimas han dejado de ser un tema de preocupación, puesto que sus mercados han evolucionado de las transacciones manuales llevadas a cabo en nombre de los inversores por máquinas de software que comercian en latencias de milisegundos (Ruta, Ruta, & Cen, 2017), inclusive se ha comprobado que gracias a la implementación de plataformas de negociación electrónica, el tiempo de negociación pueden medirse en nanosegundos (Lechman & Marszk, 2017).

Sin embargo, pese a que el comercio algorítmico es considerado una de las tendencias más importantes en el sector financiero (Brownlees, Cipollini & Gallo, 2010), en numerosos países aún no se han adoptado estas estrategias, por lo que continúan implementando exclusivamente transacciones manuales, quedando atrás frente a mercados financieros de países que cuentan con sistemas de negociación automáticos responsables de un alto volumen de negociación.

Particularmente, dentro del mercado bursátil colombiano actualmente no se cuenta con experiencias documentadas en el desarrollo de sistemas automáticos de trading; igualmente, la investigación sobre comercio algorítmico en Colombia es escasa, lo que genera un atraso del mercado colombiano y lo conduce a seguir lidiando con problemas asociados a las transacciones manuales debido a la alta intervención humana. Lo anterior incentiva que países emergentes como Colombia continúen ampliando la brecha en términos financieros respecto a países desarrollados como Estados Unidos. En este

sentido, es conveniente que los países emergentes le apuesten a estrategias automáticas de negociación, y especialmente para el caso colombiano al contar con un mercado financiero pequeño y poco líquido (Pino, Uribe & Jimenez, 2018).

De acuerdo con lo anterior, a pesar de los múltiples beneficios que genera el AT para los mercados financieros, aún su participación en numerosos países como Colombia es escasa. Se ha demostrado que el comercio algorítmico disminuye los errores humanos en el proceso de negociación, incrementa el volumen negociado a bajas latencias, aumenta la eficiencia operativa y la liquidez (Seo & Chai, 2013) y trata de minimizar los costos de transacción (Brownlees, Cipollini & Gallo, 2010), por lo que resulta valioso e importante que los mercados adopten esta estrategia. Adicionalmente, se ha manifestado que dichos beneficios son particularmente importantes en el caso de las economías en desarrollo donde los costos financieros de las transacciones suelen ser más altos y la liquidez y la eficiencia de los precios son más bajas que en las economías avanzadas (Blitz & Huij, 2012).

Es importante resaltar que el problema que posee el mercado financiero colombiano al depender mayoritariamente de las transacciones humanas no se soluciona replicando exactamente las estrategias algorítmicas que han llevado a cabo otros mercados referentes o pioneros en el tema, ya que cada mercado se comporta diferente al tener unas características y condiciones distintas. A causa de esto, no basta únicamente con implementar, sino que es necesario realizar previamente investigaciones sobre comercio algorítmico enfocadas en el mercado financiero colombiano, por lo que es significativo estudiar que se ha hecho en otros mercados, cómo se ha hecho, cómo se puede adaptar y qué aspectos se pueden incluir para implementar un sistema algorítmico de negociación en el mercado financiero colombiano.

Adicionalmente, dado que el mercado de valores de Colombia ha tenido un crecimiento significativo en los últimos años, esto ha permitido el desarrollo de otro mercado importante como es el de divisas y la integración con las bolsas de otros países, conformando el Mercado Integrado Latinoamericano (MILA) (Villada, Muñoz & García, 2012). De esta manera, para continuar con un crecimiento sostenido en el mercado de valores de Colombia es importante aplicar otras técnicas que generen valor, eficiencias e integraciones en el mercado, por lo cual el AT se convierte en un gran aliado para contribuir

con la optimización y crecimiento del mercado de valores. Así pues, la adopción de las TIC dentro de los mercados financieros incentiva la atracción de inversores extranjeros y contribuye a su integración en el sistema financiero mundial (Lechman & Marszk, 2017).

Finalmente, con la realización de esta investigación se pretende tener un acercamiento al trading automático en Colombia, particularmente a través de la aplicación de técnicas de inteligencia artificial, ya que como se mencionó anteriormente, la IA se ha posicionado fuertemente en el campo del Comercio Algorítmico, en donde se ha encontrado que un alto porcentaje de estudios muestra que los modelos de inteligencia artificial producen predicciones financieras razonablemente precisas (Barbosa & Belo, 2008).

De esta manera, en esta investigación se dará respuesta el interrogante: *¿cuáles metodologías de Inteligencia Artificial son óptimas para aplicar en el mercado financiero colombiano con el fin de generar algoritmos de negociación que generen rentabilidad?* Así, el desarrollo de este estudio contribuye a ampliar el espectro investigativo de AT en Colombia, a la vez que pretende proponer un modelo de simulación construido a partir de técnicas de IA que permita especificar estrategias de inversión y nivel de riesgo aceptado con el objetivo de que automáticamente se elija de manera óptima un portafolio de inversiones.

1.2 Revisión de Literatura

Como se mencionó anteriormente, pese que en Colombia ha sido escasa la investigación sobre comercio automático, a nivel internacional se han realizado una diversidad de investigaciones en torno a esta temática, en donde se han abordado diferentes métodos para construir algoritmos de negociación, por lo que resulta importante realizar una revisión de la producción científica de este campo de investigación. Se ha señalado que el interés de AT entre los investigadores se ha incrementado, acelerando el número de estudios publicados sobre comercio automático desde sus inicios hasta la actualidad.

En este orden de ideas, se realizó un mapeo sistemático de la literatura (SLM) con el fin de obtener una visión sintetizada de la investigación llevada a cabo en el área bajo estudio (Pellens, da Silva & da Silva, 2017). La revisión se lleva a cabo a partir de estudios

publicados sobre AT, exponiendo su avance en términos de productividad (cantidad) y el impacto (calidad) que han generado dichas publicaciones en la comunidad científica. Esto se desarrolla a través de un diagnóstico de la evolución y el estado actual de la temática, conociendo la dinámica de la publicación anual, los principales autores, revistas y países que han publicado sobre el tema, a la vez se pretende mapear los activos y técnicas más utilizadas y estudiadas sobre AT.

La investigación presentada es particularmente relevante tanto para los investigadores y profesionales de las ciencias de la computación y de la analítica debido a los retos inherentes en AT, así como el impacto de este tipo de metodologías tanto para las unidades de tesorería y mesa de negocios de las organizaciones, como para los agentes que ejercen la vigilancia y regulación del mercado. Especialmente, este es un área que ha sido poco abordada desde el naciente campo de la analítica moderna, en donde aún reposan muchos problemas por resolver.

1.2.1 Metodología

La metodología empleada en esta revisión está basada en un mapeo sistemático de la literatura (MSL), el cual permite obtener una visión sintetizada de la investigación llevada a cabo en el área bajo estudio (Pellens, da Silva & da Silva, 2017). Se ha señalado que el MSL es un tipo de estudio secundario que tiene como objetivo caracterizar un área particular de la investigación a través de un procedimiento sistemático cuyo objetivo es identificar el alcance y la naturaleza de los estudios primarios disponibles en el área (Petersen, Feldt, Mujtaba & Mattsson, 2008). A continuación, se describe el protocolo de investigación utilizado.

1.2.1.1 Proceso de búsqueda

Para la búsqueda de información bibliográfica se seleccionó la base de datos Scopus puesto que ofrece acceso a diferentes bases de datos interdisciplinarias (Mirabovich, 2014). La cadena de búsqueda utilizada fue construida a partir de las palabras clave suministradas en artículos consultados en la preparación inicial de la investigación. La cadena de búsqueda utilizada fue la siguiente:

```
( TITLE ( ( automat* OR algorithm* OR "high frequency" OR "high-  
frequency" ) PRE/1 trading ) ) OR ( KEY ( ( automat* OR  
algorithm* OR "high frequency" OR "high-frequency" ) PRE/1  
trading ) )
```

Para la búsqueda no se realizó ninguna restricción de período de tiempo, esto con el fin de que la ecuación arrojara, según los registros de Scopus, los primeros documentos que fueron publicados con respecto al AT. La investigación fue realizada en septiembre 2020.

1.2.1.2 Preguntas a resolver con el MSL

El objetivo de este mapeo es responder las siguientes preguntas de investigación:

- RQ1.** ¿Cuál es la tendencia en el número de publicaciones anuales sobre AT?
- RQ2.** ¿Cuáles son los principales autores que más han publicado sobre el tema?
- RQ3.** ¿Cuál es la dinámica de publicación de los autores?
- RQ4.** ¿Cuáles son las principales revistas que han publicado sobre el tema?
- RQ5.** ¿Cuáles son las revistas que tienen el mayor número de citas (impacto) de artículos sobre AT?
- RQ6.** ¿Cuáles son las publicaciones más citadas?
- RQ7.** ¿Cuáles son los países donde más se han concentrado las publicaciones sobre comercio automático?
- RQ8.** ¿Cuáles son los tipos de activos financieros más estudiados?
- RQ9.** ¿Cuáles son los métodos más utilizados en las investigaciones sobre AT?
- RQ10.** ¿Cuáles son las palabras clave más utilizadas en las publicaciones sobre comercio algorítmico?

Para dar respuesta a las anteriores preguntas de investigación, se consolidó una base de datos con 1.116 publicaciones arrojadas por la cadena de búsqueda, lo cual permitió realizar el respectivo mapeo. En este sentido, este estudio será de gran utilidad para que la comunidad científica conozca el alcance e impacto que han tenido las investigaciones realizadas sobre AT, a la vez que les servirá de insumo para establecer futuras líneas de investigación. Por otra parte, este mapeo también será de utilidad a los profesionales de diferentes áreas para establecer las técnicas más usadas que han sido aplicadas para construir algoritmos de negociación.

1.2.2 Resultados obtenidos

Al buscar en la base de datos de Scopus, se recuperaron automáticamente 1.176 documentos a partir de la cadena de búsqueda expuesta anteriormente. Luego se aplicaron manualmente los criterios de inclusión y exclusión; como resultado, finalmente se seleccionaron 1.116 documentos, correspondientes a 635 artículos en revistas, 404 artículos de conferencia, 68 capítulos de libros, 8 libros y 1 artículo en proceso de impresión. Los 1.116 artículos fueron publicados en 595 fuentes diferentes entre 1984-2021, los cuales fueron escritos por un total de 2.088 autores afiliados a 891 instituciones en 65 países. El 78% de los artículos fueron escritos por dos o más autores, mientras que el 22% restante son artículos de un solo autor, resultando una media de 1,87 autores por documento y 0,53 documentos por autor. En cuanto al impacto, cada artículo tiene una media de 8,13 citas.

Es importante tener en cuenta que la ecuación de búsqueda arrojó registros publicados desde los inicios de la temática hasta septiembre 2020 que se realizó la búsqueda, por lo cual no se cuenta con los documentos completos del año 2020 ya que para ese año es posible que existieran registros sometidos a procesos de evaluación, revisión e impresión en revistas indexadas en Scopus. También en esta base de datos, se encuentran artículos publicados disponibles a septiembre de 2020 pero con una fecha de publicación posterior.

1.2.3 Respuestas a las preguntas de investigación

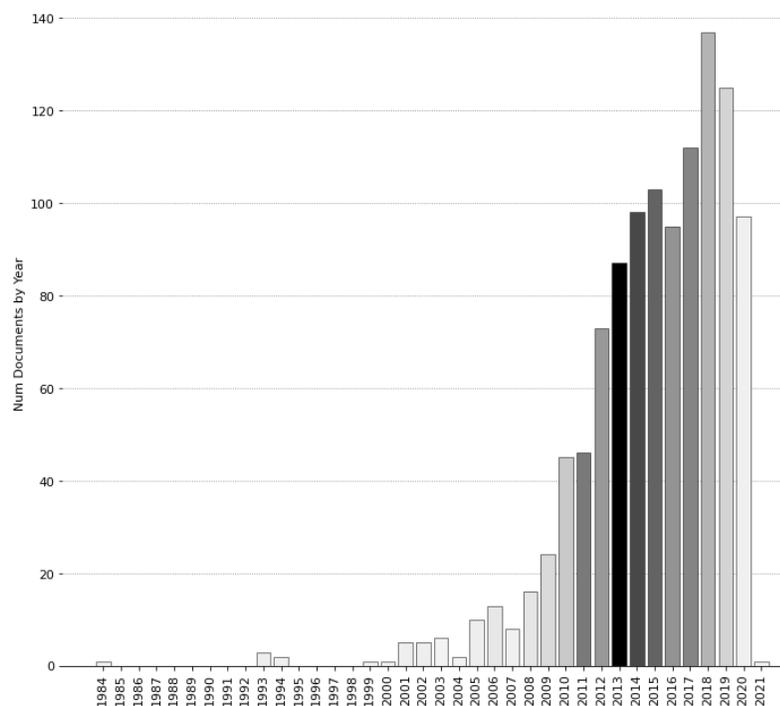
En esta sección se dará respuesta a las preguntas de investigación planteadas:

RQ1. ¿Cuál es la tendencia en el número de publicaciones anuales sobre AT?

La Figura 1-1 presenta el número de documentos publicados por año durante el período observado. La figura muestra una tendencia ascendente, revelando el creciente interés de los investigadores y la relevancia de este tema de investigación. La tasa de crecimiento de las publicaciones es del 20,28% anual durante los últimos cinco años. Esta tendencia sigue la Ley Bibliométrica de Crecimiento Exponencial, que postula que la producción científica existente se duplica cada 10 o 15 años (Ardanuy, 2012).

El número máximo de publicaciones se da en el año 2018 con un total de 137 artículos. Sin embargo, 2018 no es el año con mayor número de citas, ya que la intensidad del color de las barras representa el número de citas. En este sentido, aunque años como 2013 y 2014 no tienen tantos artículos publicados, han generado más impacto, ya que los artículos de estos años han sido citados muchas más veces en comparación con años de mayor número de publicaciones.

Figura 1-1: Publicaciones por año.



Nota: para las siguientes figuras, el primer número después del nombre de cada autor, fuente, activo, palabra clave o técnica de IA indica el número de publicaciones asociadas a ese elemento, mientras que el segundo número indica el número total de citas.

RQ2. ¿Cuáles son los principales autores que más han publicado sobre el tema?

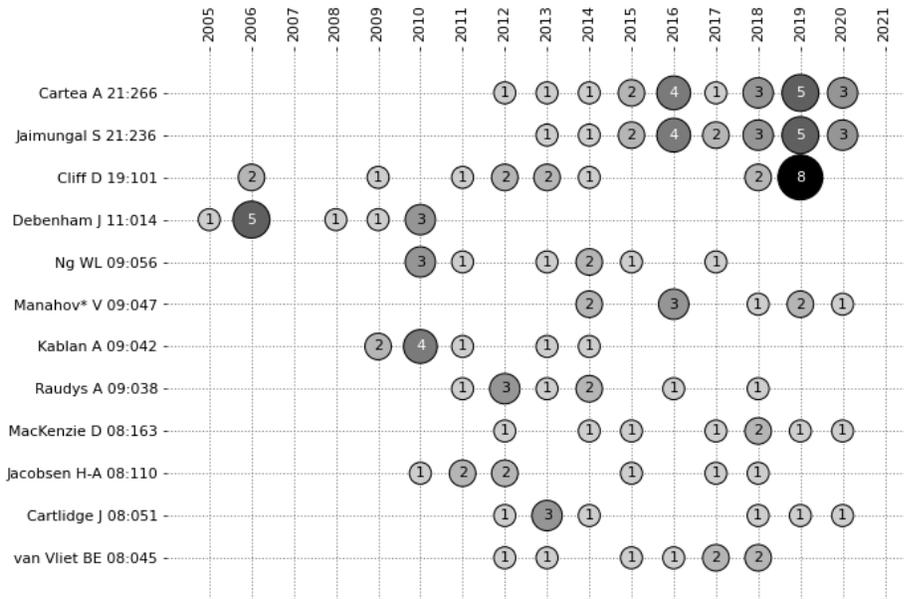
Los 1.116 documentos seleccionados fueron escritos por 2.088 autores. La Figura 1-2 muestra los 12 autores con más publicaciones sobre AT; donde se observa que Cartea y

Jaimungal son los autores más importantes con 21 documentos cada uno, seguidos por Cliff con 13 documentos. Los dos autores principales han compartido coautoría en varios artículos, lo cual se verá con más detalle en la siguiente pregunta.

Además, se puede observar que Cartea y Jaimungal se han mantenido vigentes en el campo desde su primera publicación (2012 y 2013 respectivamente); estos autores han seguido publicando en todos los años hasta 2020. En uno de sus estudios, Cartea y Jaimungal proponen un modelo en el que un operador algorítmico analiza la distribución de precios en una fecha futura y luego decide cómo operar. Ambos autores muestran cómo este trader (operador) algorítmico obtiene mejores rendimientos que un operador común, ya que el primero puede aprender de la dinámica del mercado y actualizar sus predicciones para ajustar su estrategia (Cartea, Jaimungal & Kinzebulatov, 2016).

En este punto cabe señalar que, aunque Kissell (2013) no se encuentra entre los autores principales, se le considera el primer autor en analizar el trading algorítmico en las diferentes clases de activos, proporcionando información valiosa para los traders para desarrollar y construir sus propios algoritmos de negociación.

Figura 1-2: Principales autores de AT con su número de publicaciones por año.

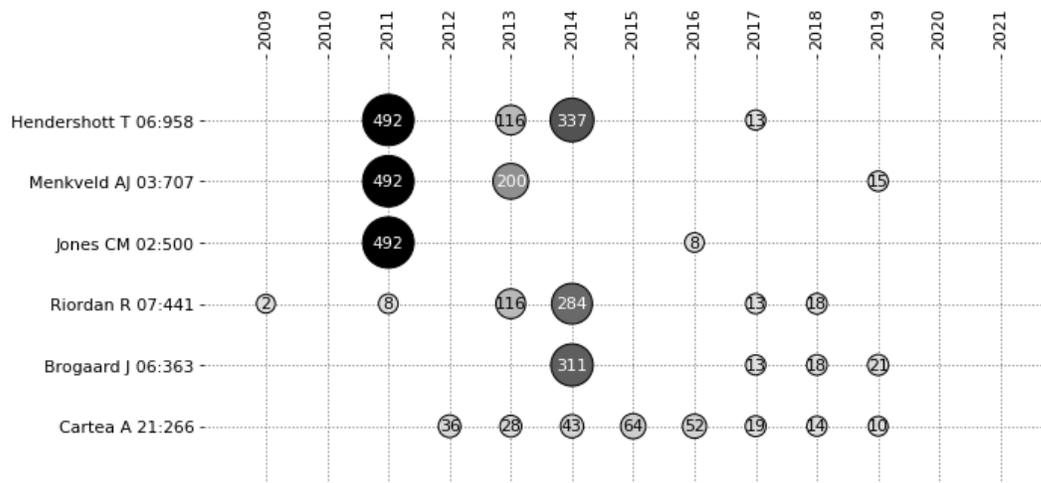


Por otro lado, aunque Hendershott no se encuentra entre los autores que han publicado mayor cantidad de artículos, si se encuentra entre los autores con mayor cantidad de

citaciones, con un total de 958 citas y 6 artículos publicados (ver Figura 1-3). Entre los temas abordados por este autor se encuentra el efecto que tiene el trading algorítmico en la liquidez (Hendershott, Jones & Menkveld, 2011; Brogaard, Hendershott & Riordan, 2014; Brogaard, Hendershott & Riordan, 2017; Brogaard, Hendershott, Hunt & Ysusi, 2014; Ding, Hanna & Hendershott, 2014; Hendershott & Riordan, 2013). El segundo autor más citado es Menkveld (Hendershott, Jones & Menkveld, 2011; Van Kervel & Menkveld, 2019; Menkveld, 2013), quien está lejos de estar en el top de los principales autores ya que solo cuenta con 3 documentos, pero ocupa el lugar número 2 en autores con mayor impacto con un total de 707 citas. Más de la mitad de las citas de cada uno de los dos autores anteriores provienen de un solo artículo en el que comparten coautoría: *Does algorithmic trading improve liquidity?* (2011) (Hendershott, Jones & Menkveld, 2011) con 492 citas, que como se verá más adelante, es el artículo más citado de acuerdo a esta investigación.

Finalmente, Cartera, que es uno de los dos autores con mayor número de artículos publicados, ocupa el puesto número 6 entre los autores con mayor número de citas (impacto).

Figura 1-3: Cantidad de citas por año para los 6 autores más citados.



RQ3. ¿Cuál es la dinámica de publicación de los autores?

La Tabla 1-1 presenta el número de documentos escritos por un mismo autor. La tabla indica que 1.706 autores han escrito solo un artículo, 240 autores han escrito dos artículos y así sucesivamente. Los resultados muestran que el 82% (1.706 / 2.088) de los autores ha publicado un solo artículo, el 11% ha publicado dos documentos y solo el 7% restante ha escrito tres o más artículos. Estos resultados señalan que no hay muchos expertos en la temática, ya que la mayoría de los autores son fugaces y por tanto son pocos los que continúan publicando y madurando el campo de la investigación. Este elevado número de autores con una única publicación indica un alto porcentaje de autores esporádicos, quienes exploran este campo y posiblemente migran a otros temas o actividades (Urbizagástegui & Cortés, 2002).

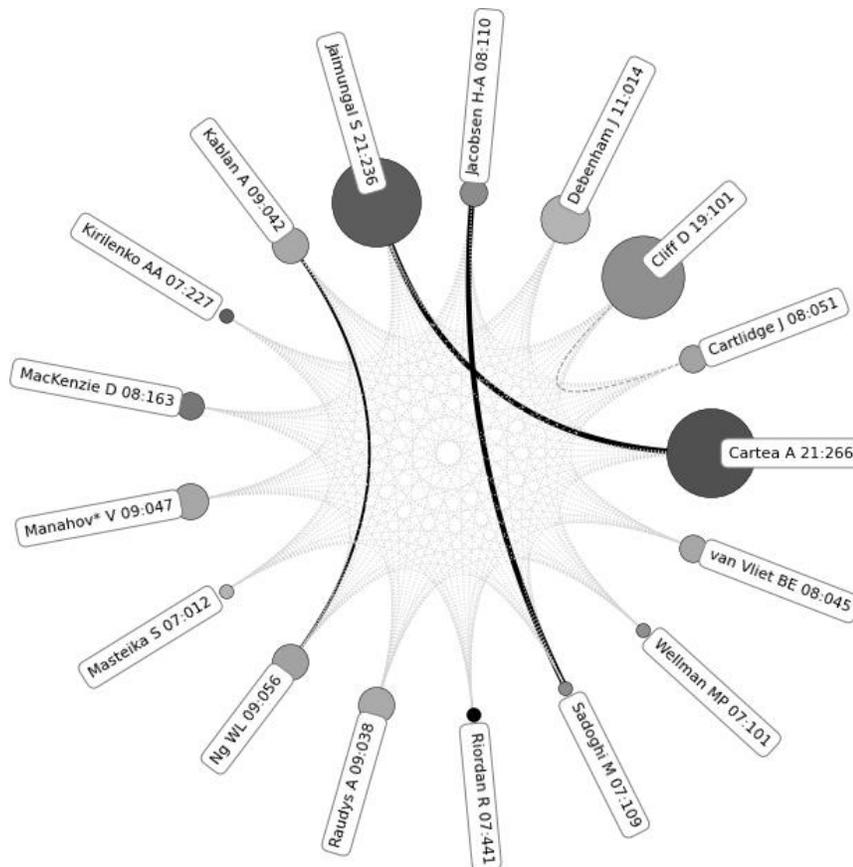
Tabla 1-1: Número total de documentos escritos por un mismo autor.

Número de documentos escritos por autor	Total de autores
1	1.706
2	240
3	65
4	32
5	17
6	11
7	5
8	4
9	4
11	1
19	1
21	2
Total	2.088

Por su parte, la Figura 1-4 muestra la relación de coautoría entre los 17 autores con más artículos publicados sobre AT; dichos autores han publicado 7 o más documentos. Por un lado, Cartea (Reino Unido) y Jaimungal (Canadá) son los dos autores con mayor número de artículos (21 cada uno) y comparten la coautoría en 17 de ellos. Por otro lado, Kablan (Malta) y Ng WL (Reino Unido) han publicado 9 documentos cada uno y comparten la coautoría en 5 de ellos. Finalmente, Sadoghi (Canadá) ha publicado 7 artículos, todos ellos en coautoría con Jacobsen (Canadá). Como se puede ver, es común que autores de

diferentes países comparten coautoría; en este contexto, se ha definido que la coautoría internacional es uno de los medios de colaboración hacia la internacionalización (Ng, Mustafa & Kurian, 2019).

Figura 1-4: Coautoría entre autores principales.



RQ4. ¿Cuáles son las principales revistas que han publicado sobre el tema?

Los 1.116 artículos seleccionados se publicaron en 595 fuentes. El 73% de las revistas tiene una sola publicación, lo que refleja que no hay muchas revistas especializadas en el tema; el 13% de las revistas tiene dos publicaciones y el 14% restante tiene 3 o más documentos publicados. De acuerdo con la Ley de Bradford, las revistas con 6 o más documentos (27 revistas) representan las revistas principales que han publicado sobre AT. En la Figura 1-5 se presenta la cantidad de publicaciones por año para las 11 revistas con

mayor número de artículos publicados sobre AT (11 documentos o más por fuente), las cuales cubren el 19% de los artículos seleccionados.

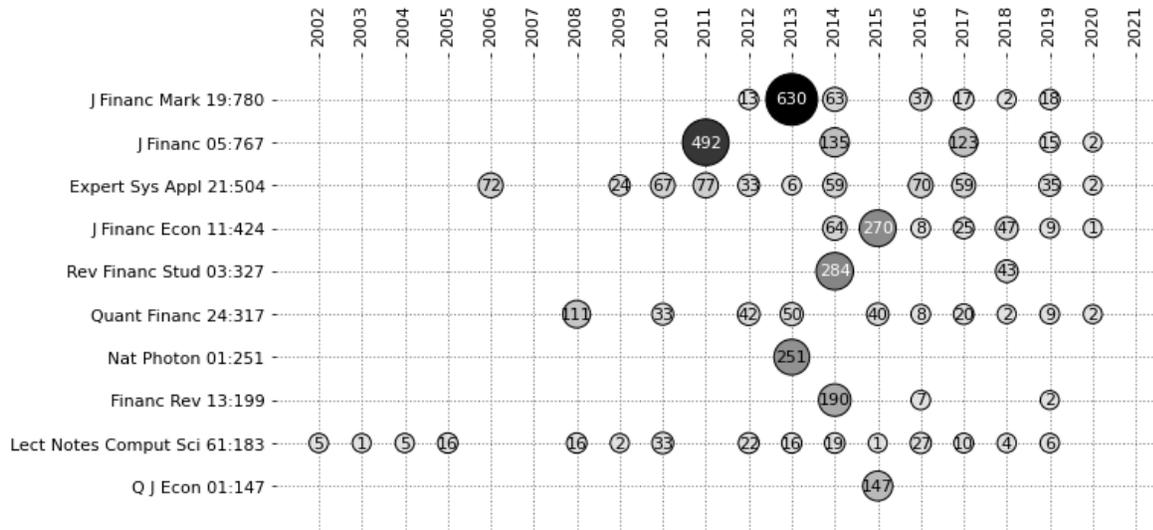
Figura 1-5: Publicaciones por año de las principales revistas.

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
Lect Notes Comput Sci 61:183	0	0	0	2	1	1	2	2	0	2	1	9	2	6	3	7	1	3	7	3	7	2	0	
Quant Financ 24:317	0	0	0	0	0	0	0	0	0	1	0	1	0	3	1	0	3	2	4	2	3	4	0	
Expert Sys Appl 21:504	0	0	0	0	0	0	0	1	0	0	1	1	3	2	1	1	0	3	3	0	2	2	1	
Handb of High Freq Trading 20:030	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	
J Financ Mark 19:780	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	3	0	3	2	1	4	1	0	
Financ Rev 13:199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	1	0	0	2	1	0	
Res Int Bus Financ 13:029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	0	0	6	0	
Phys A Stat Mech Appl 12:060	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	1	1	2	2	1	0	
J Financ Econ 11:424	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	2	2	1	1	0	
Finan Res Lett 11:052	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	4	0	
ACM Int Conf Proc Ser 11:009	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	3	3	0	0

La revista con más artículos es *Lecture Notes in Computer Science* (Alemania), con 61 documentos entre 2002 y 2020; sin embargo, la importancia de esta revista es muy baja ya que la cantidad de citas por artículo es 3,0; en comparación con otras revistas como el *Journal of Finance* con 153,4 citas por artículo.

RQ5. ¿Cuáles son las revistas que tienen el mayor número de citaciones (impacto) de artículos sobre AT?

La Figura 1-6 presenta las 10 revistas con mayor número de citaciones. La primera es *Journal of Financial Markets* de Holanda con un total de 780 citas obtenidas por 19 artículos, lo cual la ubica en la quinta posición de revistas con mayor cantidad de publicaciones. La segunda es el *Journal of Finance* del Reino Unido con un total de 767 citas derivadas de 5 artículos. Cabe señalar que las dos revistas con mayor número de publicaciones no se encuentran en el top 5 de las revistas con mayor impacto (citas).

Figura 1-6: Citaciones por año para las 10 fuentes más citadas.**RQ6. ¿Cuáles son los artículos más citados?**

La Tabla 1-2 presenta los 10 artículos más citados. Se encontró que el 32% de los artículos no tienen citas. El artículo más influyente es: *Does algorithmic trading improve liquidity?* (2011) (Hendershott, Jones & Menkveld, 2011). citado 492 veces, seguido de: *High-frequency trading and price discovery (2014)* (Brogaard, Hendershott & Riordan, 2014) con 284 citas. Los autores del primer artículo son Hendershott, Menkveld y Jones, quienes no se encuentran entre los 10 autores que más han publicado sobre el tema, aunque estos 3 autores representan los 3 autores con mayor impacto, dado su elevado número de citas, impulsado por escribir el artículo más citado sobre AT. En *Does algorithmic trading improve liquidity?*, sus autores, utilizando datos de la Bolsa de Valores de Nueva York, concluyen que el comercio algorítmico mejora la liquidez de las cotizaciones.

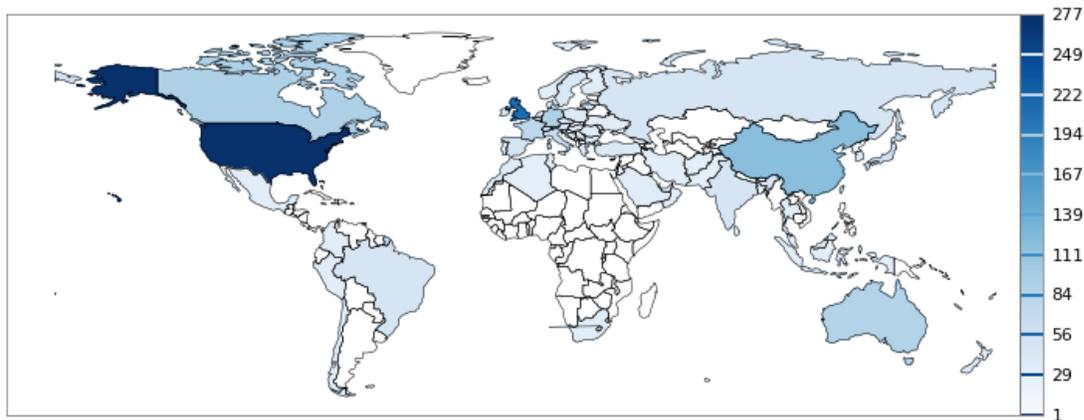
Tabla 1-2: Artículos más citados.

Título de la publicación	Autores	Fuente	Número de citaciones
Does algorithmic trading improve liquidity? (2011)	Hendershott T Jones CM Menkveld AJ	Journal of Finance	492
High-frequency trading and price discovery (2014)	Brogaard J Hendershott T Riordan R	Review of Financial Studies	284
Towards high-capacity fibre-optic communications at the speed of light in vacuum (2013)	Poletti F Wheeler NV Petrovich MN Baddela N Numkam Fokoua E Hayes JR Gray DR Li Z Slavik R Richardson DJ	Nature Photonics	251
Low-latency trading (2013)	Hasbrouck J Saar G	Journal of Financial Markets	215
High frequency trading and the new market makers (2013)	Menkveld AJ	Journal of Financial Markets	200
The high-frequency trading arms race: Frequent batch auctions as a market design response (2015)	Budish E Cramton P Shim J	Quarterly Journal of Economics	147
Rise of the machines: Algorithmic trading in the foreign exchange market (2014)	Chaboud AP Chiquoine B Hjalmarsson E Vega C	Journal of Finance	135
High frequency market microstructure (2015)	O'Hara M	Journal of Financial Economics	127
The Flash Crash: High-Frequency (2017) Trading in an Electronic Market	Kirilenko AA Kyle AS Samadi M Tuzun T	Journal of Finance	123
Algorithmic trading and the market for liquidity	Hendershott T Riordan R	Journal of Financial and Quantitative Analysis	116

RQ7. ¿Cuáles son los países donde más se ha concentrado la investigación sobre comercio automático?

La Figura 1-7 muestra la distribución de documentos en todo el mundo, donde se observa que en todos los continentes existe al menos una publicación sobre AT. Estados Unidos tiene la producción más alta con 277 documentos, lo que representa una cuarta parte de la producción académica total. El segundo es Reino Unido con 205 artículos (18%). En particular, Estados Unidos le ha dado gran importancia al comercio automático, donde estas herramientas son responsables de más del 70% del volumen de negociación de activos (Cartea & Jaimungal, 2013). Además, los principales países en publicar sobre comercio algorítmico se clasifican como países desarrollados. En este sentido, hay una falta de investigación sobre la aplicabilidad de AT en economías emergentes.

Figura 1-7: Publicaciones por país.



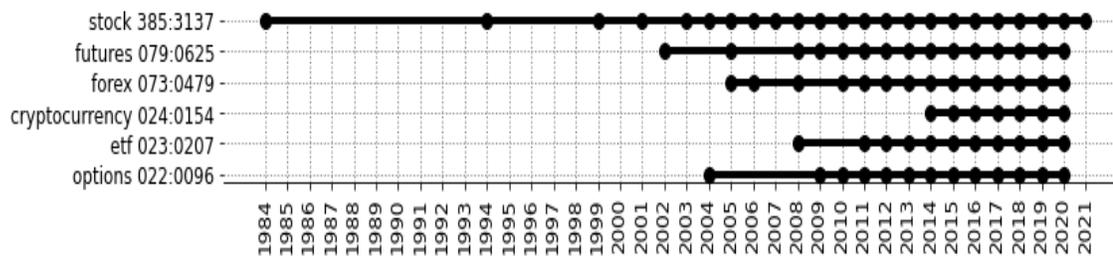
RQ8. ¿Cuáles son los tipos de activos financieros más estudiados?

La Figura 1-8 muestra el número de documentos publicados sobre comercio algorítmico por tipo de activo. Es importante tener en cuenta que una investigación puede cubrir uno o más activos. La investigación sobre el comercio automático se ha centrado en gran medida en las acciones; desde sus inicios hasta el presente, el mercado de valores se ha mantenido fuerte. En segundo lugar, se encuentran los futuros y de tercero están las divisas; este último es el mercado más grande del mundo debido a su alto volumen diario de operaciones (Roledene, Ariyathilaka, Liyanage, Lakmal & Bamunusinghe, 2016) y en la

literatura se ha mencionado que en los últimos años el mercado de divisas ha experimentado un aumento gracias a la presencia de comercio automatizado (Fukuma & Kadogawa, 2020).

Finalmente, los estudios sobre criptomonedas tuvieron sus inicios en 2014; sin embargo, la investigación sobre este tipo de activos se ha vuelto más relevante en los últimos años, dado que se ha posicionado como una opción potencial para diversificar el riesgo financiero de invertir en otros activos (Chu, Chan & Zhang, 2020).

Figura 1-8: Publicaciones por tipo de activo.



RQ9. ¿Cuáles son los métodos más utilizados en las investigaciones sobre AT?

La Figura 1-9 muestra el número de documentos por técnica implementada. Esta figura muestra que la inteligencia artificial y las técnicas de aprendizaje automático son las más utilizadas, siendo las redes neuronales artificiales (ANN) la técnica principal. En particular, se destacan técnicas como aprendizaje de Hebb, combinadores lineales adaptativos y las redes de propagación directa (Janosek, Volna & Kotyrba, 2015). Recientemente, se han comenzado a utilizar redes de aprendizaje profundo (le Calvez & Cliff, 2018).

El aprendizaje por refuerzo también se posiciona como una de las técnicas más relevantes. Es importante mencionar que en 1996 Moody y Wu propusieron el modelo Aprendizaje de Refuerzo Recurrente (RRL por sus siglas en inglés) como un método viable de aprendizaje automático dentro del comercio algorítmico utilizado para descubrir estrategias comerciales (Gabrielsson & Johansson, 2015). Asimismo, se han utilizado diferentes variantes de algoritmos evolutivos (EA) como los algoritmos genéticos (GA), la programación genética (GP) y la evolución gramatical. Otra tendencia emergente es

complementar la información utilizada en los modelos con otros datos como noticias en línea y redes sociales (Hogenboom, 2012), lo que ha permitido que los sistemas de AT reaccionen rápidamente a eventos del mercado.

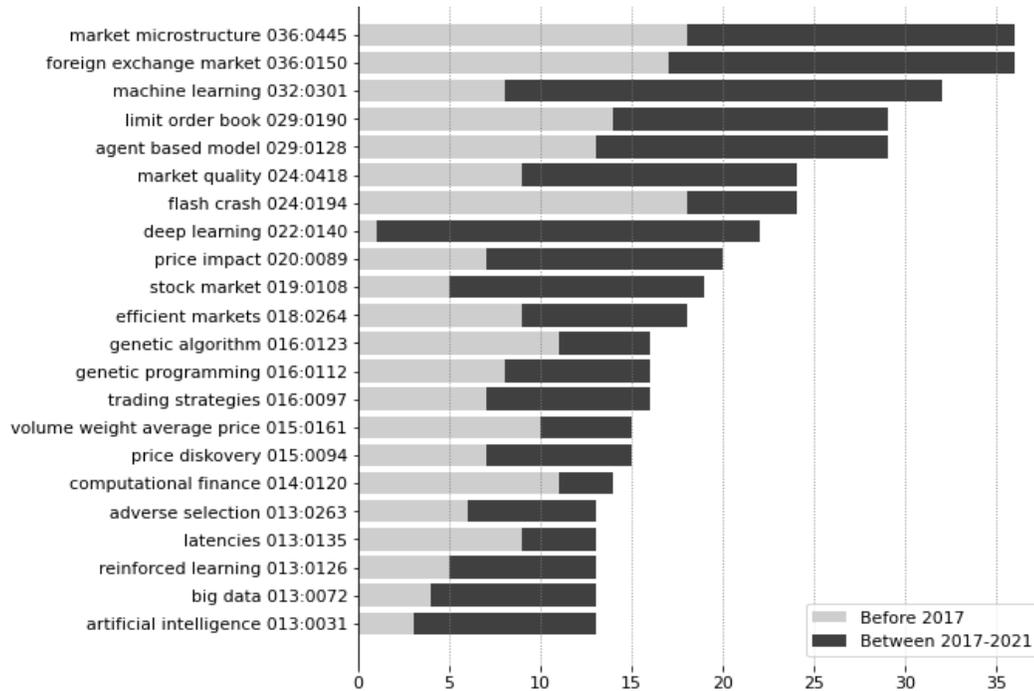
Figura 1-9: Cantidad de artículos publicados por año para las principales técnicas.

	1993	1994	1999	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
artificial neural network 71:484	1	1	1	0	1	1	1	2	2	1	2	3	10	4	3	6	18	13	1	0
reinforcement learning 32:272	0	0	0	2	1	0	0	0	2	3	2	2	4	2	1	0	2	8	2	1
genetic algorithm 29:207	0	0	1	2	3	2	0	2	3	1	1	3	2	0	3	6	0	0	0	0
moving average techniques 26:170	0	0	0	0	0	0	1	1	1	2	1	2	4	2	1	1	2	5	3	0
clustering 26:128	0	0	0	0	0	0	0	1	1	0	0	2	5	1	4	0	3	4	5	0
genetic programming 24:113	0	0	0	0	0	0	0	0	1	2	1	1	5	0	4	1	6	2	1	0
deep learning 22:153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	8	9	2	0
agent based model 17:076	0	0	0	0	0	0	0	0	0	0	0	2	0	2	2	1	3	4	3	0
lstm neural network 12:026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	8	1	0
fuzzy inference system 11:137	0	0	0	0	0	0	0	2	0	4	0	0	3	0	1	0	0	0	1	0
convolutional neural network 10:107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	2	1	0
support vector machine 10:037	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	3	3	0	0
q learning 09:044	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	4	1	1
fuzzy logic 09:035	0	0	0	0	0	0	0	0	3	1	0	1	2	1	1	0	0	0	0	0

RQ10. ¿Cuáles son las palabras clave más utilizadas en las publicaciones sobre comercio algorítmico?

Para el análisis de palabras clave se realizó una agrupación manual de términos que difieren en la escritura pero que tienen el mismo significado o apuntan al mismo concepto. Tras esta limpieza se obtuvieron un total de 2.112 palabras clave diferentes. La Figura 1-10 muestra las palabras clave más frecuentes y compara su frecuencia en 2 periodos: antes de 2017 vs apariciones entre 2017 y 2021. Por ejemplo, el aprendizaje profundo es un tema nuevo en las publicaciones de AT, ya que casi todas sus menciones se dan entre los años 2017 y 2021, solo una parte muy pequeña fue antes de 2017. Algo similar ocurre con los términos *machine learning* e *inteligencia artificial*, lo cuáles han tomado mucha más fuerza en los últimos cinco años, debido a que son técnicas altamente utilizadas para crear algoritmos de AT (Berutich, López, Luna & Quintana, 2016; Ta, Liu & Addis, 2018).

Figura 1-10: Cantidad de apariciones de palabras clave en artículos sobre AT antes de 2017 vs 2017-2021.



RQ11. ¿Cuáles son las principales temáticas abordadas en AT?

Para encontrar las principales temáticas abordadas, se realizó el agrupamiento de términos de la matriz de co-ocurrencias de palabras clave, el cual permite encontrar relaciones o interacciones entre palabras. El procedimiento realizado fue el siguiente:

- Se construyó la matriz de co-ocurrencias para las palabras clave con una ocurrencia mínima de 20.
- Se realizó la normalización de la matriz de co-ocurrencias utilizando el índice de asociación.
- Se construyeron clústeres utilizando el algoritmo de agrupamiento de Louvaine.

Mediante la aplicación de la metodología descrita se obtuvo un total de 3 grupos temáticos que son presentados en la Tabla 1-3. Al analizar el primer clúster resultante, se observa que las palabras contenidas en él se refieren a inteligencia artificial, aprendizaje automático

y sus diferentes técnicas. Particularmente en este clúster cobran importancia las redes neuronales artificiales y sus diversos tipos, donde se destacan *LSTM Neural Network* y *Convolutional Neural Network*. En este punto es importante resaltar que, si bien se sabe que el aprendizaje profundo se ha aplicado con éxito en el procesamiento de imágenes, el reconocimiento de voz, la traducción mecánica y otros campos, varias investigaciones de AT han demostrado cómo la aplicación de algoritmos de aprendizaje profundo en el comercio de alta frecuencia genera mejores resultados en comparación con las estrategias comerciales tradicionales (Fister, Mun, Jagrič & Jagrič, 2019). Adicionalmente, el clúster 1 se complementa con los diferentes activos de los mercados financieros como Forex (mercado de divisas), acciones (mercado de capitales - renta variable) y criptomonedas. Esta co-ocurrencia ocurre porque las técnicas de IA son un medio potente para generar señales de compra y venta (AT) de este tipo de activos.

Por otro lado, en el clúster 2 hay palabras relacionadas con la microestructura, eficiencia y liquidez del mercado. Por ejemplo, en este grupo se encuentra la palabra "calidad del mercado", que se mide por la liquidez y la eficiencia (Veryzhenko, Arena, Harb & Oriol (2017), y por lo tanto está estrechamente relacionada con palabras de este grupo como "mercados líquidos" y "mercados eficientes".

Finalmente, en el clúster 3, con menos palabras, se hace referencia a temas de portafolio. Particularmente, la co-ocurrencia en este grupo entre "arbitraje estadístico" e "hipótesis de mercado eficiente" ocurre porque algunas investigaciones encontraron que el arbitraje estadístico en ambiente de comercio de alta frecuencia (HFT por sus siglas en inglés) refuta la hipótesis de mercado eficiente (Driaunys, Masteika, Sakalauskas & Vaitonis, 2014; Masteika & Vaitonis, 2015).

Tabla 1-3: Grupo de términos obtenidos de la matriz de concurrencia de palabras clave

Clúster	Palabras clave	
1	<ul style="list-style-type: none"> • FOREIGN EXCHANGE MARKET • MACHINE LEARNING • ARTIFICIAL NEURAL NETWORK • TECHNICAL ANALYSIS • DEEP LEARNING • STOCK MARKET • GENETIC ALGORITHM • GENETIC PROGRAMMING • TRADING STRATEGIES • COMPUTATIONAL FINANCE • REINFORCED LEARNING 	<ul style="list-style-type: none"> • BIG DATA • ARTIFICIAL INTELLIGENCE • FINANCIAL FORECASTING • MOVING AVERAGE TECHNIQUES • BITCOIN • CRYPTOCURRENCIES • DIRECTIONAL CHANGE • SUPPORT VECTOR MACHINE • LSTM NEURAL NETWORK • CONVOLUTIONAL NEURAL NETWORK
2	<ul style="list-style-type: none"> • MARKET MICROSTRUCTURE • LIMIT ORDER BOOK • AGENT BASED MODEL • MARKET QUALITY • FLASH CRASH • PRICE IMPACT • EFFICIENT MARKETS • MARKET MAKING • VOLUME WEIGHT AVERAGE PRICE 	<ul style="list-style-type: none"> • PRICE DISCOVERY • ADVERSE SELECTION • LATENCIES • MARKET REGULATION • STOCHASTIC CONTROL • OPTIMAL EXECUTION • LIQUID MARKETS • BID ASK SPREAD
3	<ul style="list-style-type: none"> • STATISTICAL ARBITRAGE • OPTIMAL PORTFOLIO • EFFICIENT MARKET HYPOTHESIS 	<ul style="list-style-type: none"> • MULTIAGENT SYSTEM • PAIRS TRADING • FUTURES MARKET

1.3 Discusión

La revisión de literatura realizada hasta el momento sobre Comercio Algorítmico muestra que este es un tema que está tomando cada vez más relevancia a nivel mundial, tanto en el campo investigativo como en la aplicación. En este sentido, se puede inferir que la investigación aplicada es un gran aliado del AT, dado que permite transformar el conocimiento teórico proveniente de la investigación básica en conceptos y prototipos tangibles en los mercados financieros (Lozada, 2014), lo que conduce a que éstos puedan operar de una manera más automática y eficiente.

En este sentido, se evidenció que los países que más publican sobre AT son quienes más han implementado algoritmos de negociación en sus mercados financieros, por lo que se muestra una correlación entre dichas variables (investigación y aplicación). Este es el caso de Estados Unidos (Ver Figura 1-7), el cuál es el país líder en publicar investigaciones

sobre Comercio Automático y al tiempo es destacado por sus numerosas aplicaciones en el mercado financiero, en donde más del 70% del volumen negociado en acciones se da a través del comercio automático. Asimismo, en otras regiones como Europa y Asia también existe una participación considerable de los algoritmos de negociación (Kunz & Martin, 2015).

Sin embargo, para los países emergentes este panorama no es el mismo, como se pudo ver en la Figura 1-7, estos países no sobresalen en la comunidad científica en temas relacionados con Comercio Algorítmico, por lo que su aplicación hoy en día en los mercados financieros es escasa. Esta situación es debida en gran parte a la menor tasa de penetración de las TIC que tienen las economías emergentes frente a las más avanzadas (Lechman & Marszk, 2017). Precisamente, como se mencionó en el planteamiento del problema, este escenario es el que presenta actualmente Colombia, el cual no presenta muchos avances en materia de Comercio Automático, tanto en temas investigativos como de aplicación, por lo que su mercado presenta un atraso en comparación con otros países. En este orden de ideas, existe un gran reto para los países en investigar y aplicar las diversas técnicas de AT con el fin de impulsar sus mercados a ser mucho más eficientes y competitivos.

Con respecto a los métodos para generar algoritmos de negociación, actualmente los países pioneros y referentes en el tema han estudiado y aplicado una diversidad de técnicas para la generación de AT. Se ha indicado que el pronóstico del precio (valor o dirección) resulta de alto interés para los inversionistas ya que les da señales de compra y venta de estos títulos con el fin de maximizar sus beneficios. Adicionalmente, dado que las series de tiempo de las acciones se caracterizan por la no linealidad y la discontinuidad ocasionada por los múltiples factores que interactúan el mercado bursátil, las técnicas de Inteligencia Artificial (AI) como las Redes Neuronales, la Programación Genética y la Lógica Difusa se han convertido en técnicas populares de investigación que se han estudiado e implementado para la predicción de series de tiempo (Hadavandi et al., 2010).

Específicamente, las redes neuronales representan el método más utilizado en el AT (Ver Figura 1-9). En la literatura se ha manifestado que esta técnica se ha implementado con éxito en varias aplicaciones financieras. Particularmente, en el mercado bursátil y en los

dominios de divisas, las redes neuronales se han usado con el objetivo de predecir los precios futuros de las acciones y los pares de divisas, calcular su tasa de rendimiento y analizar los riesgos (Kayal, 2010; Sezer, Ozbayoglu & Dogdu, 2017). Aunque para el caso colombiano es poca la investigación que se ha generado en torno en el Comercio Automático, en el año 2012 se realizó un estudio sobre un modelo basado en redes neuronales artificiales para el pronóstico de los precios de dos de las principales acciones transadas en mercado de valores colombiano: Ecopetrol y Preferencial Bancolombia, empresas que negocian en las bolsas de valores de Colombia y Nueva York (Villada, Muñoz & García, 2012).

Por su parte, la programación genética también se posiciona como una técnica bastante estudiada y utilizada en el Comercio Automático, la cual se ha implementado para determinar las posibles condiciones de compra y venta de acciones de acuerdo a ciertas reglas de negociación. Esta técnica es capaz de ofrecer soluciones robustas que soportan las condiciones extremas del mercado, al tiempo que produce altos rendimientos con un riesgo mínimo (Berutich, López, Luna & Quintana, 2016).

Si bien en la Figura 1-9 se evidencia la investigación y aplicación de algunas técnicas de AT en los mercados financieros, también se puede deducir que existe una amplia variedad de métodos que no han sido estudiados en detalle; por lo anterior, aún es necesario determinar cuáles son las técnicas que mejor funcionan o se adecúan al mercado financiero objetivo o cuál es la combinación óptima de técnicas de IA que podrían emplearse de acuerdo a sus diferentes fortalezas y debilidades. Esto con el fin de constituir un modelo híbrido de técnicas que generen mejores resultados que al implementarlo individualmente (Hadavandi et al., 2010). En este marco, en la comunidad científica ya se han implementado varios modelos híbridos que buscan obtener mejores resultados que al aplicar técnicas de manera individual.

En este contexto, se ha señalado que uno de los enfoques más populares es la hibridación entre la lógica difusa y programación genética que conducen a los sistemas genéticos difusos (GFS) (Cordón, Herrera, Holfmann & Magdalena, 2001). Por ejemplo, Atsalakis y Valavanis (2009) propusieron un modelo híbrido para pronosticar las tendencias de los precios de las acciones del día siguiente de Athens y New York Stock Exchange (NYSE). Por su parte, Hadavandi et al. (2010) propuso un modelo híbrido que combina algoritmos

genéticos, lógica difusa y redes neuronales para pronosticar el precio de dos acciones de tecnología y aerolíneas.

Además de seleccionar la técnica adecuada para implementar AT, otro aspecto importante en el Comercio Automático es determinar el tipo de activo sobre el cual se aplicarán los algoritmos de negociación con el fin de efectuar las transacciones (compra y venta) de manera óptima. En la comunidad científica, las acciones, futuros y FOREX resaltan como los activos más investigados, lo que es de esperarse dado que las acciones sobre empresas son los títulos más transados comúnmente en un mercado de valores, el cual a su vez es un componente fundamental de los mercados financieros (Villada, Muñoz & García, 2012). Sin embargo, para el caso colombiano es poco lo que se ha investigado sobre la eficiencia del mercado accionario (Morales, Yopez & Castañeda, 2015), lo cual representa un atraso del país frente a otros países que ya han adelantado en estudios y aplicaciones relacionadas a este tema.

Se concluye entonces que, para las economías emergentes, y particularmente para Colombia, hay mucho por trabajar en materia de AT. Una de las posibles líneas de investigación es comparar la efectividad de diversas técnicas de IA, sean aplicadas de manera individual o a partir de modelos híbridos, en la predicción de precios o movimientos de algunas acciones particulares que transen en la Bolsa de Valores de Colombia, de tal manera que sean la base para proponer algoritmos de negociación óptimos para la compra y venta de acciones.

1.4 Hipótesis

Es posible construir un modelo de simulación a partir de técnicas de Inteligencia Artificial (IA) que permita generar decisiones automáticas de compra y venta de acciones para la conformación óptima de un portafolio en la Bolsa de Valores de Colombia.

1.5 Objetivos

1.5.1 Objetivo General

Construir un prototipo de simulación para evaluar estrategias de Comercio Algorítmico en el mercado financiero colombiano, con el fin de analizar si es posible incorporar este tipo de estrategias por parte de los inversionistas.

1.5.2 Objetivos Específicos

- Proponer una estrategia para desarrollar modelos que permitan pronosticar el movimiento de precios para las principales acciones en el mercado financiero colombiano.
- Adaptar una estrategia para la construcción de portafolios que incorpore los modelos de pronósticos desarrollados en el objetivo anterior.
- Validar el modelo desarrollado en los Objetivos 1 y 2 mediante un modelo de simulación el cual permita especificar estrategias para decisiones de conformación de portafolio.

1.6 Mapa del documento

El resto de esta tesis está organizada así: en el Capítulo 2, se presenta la metodología propuesta para desarrollar la pregunta de investigación y objetivos planteados, por lo que en el Capítulo 3 se ejecuta dicha metodología donde se presentan los resultados obtenidos. Finalmente, en el Capítulo 4 se cierra este trabajo con las conclusiones.

2. Aproximación Metodológica Propuesta

En este capítulo se explica la metodología implementada para la conformación automática de portafolios de inversión, el cuál es constituido por las acciones vigentes que conforman la canasta del COLCAP. Para construir la estrategia de inversión, el primer paso es la predicción del movimiento del precio de las acciones. Esta predicción indica si el precio al día siguiente estará al alza o a la baja, lo cual da señales al inversionista del momento óptimo para abrir o cerrar posiciones. En este sentido, en la primera parte de esta metodología se explicará la manera como se llevaron a cabo las predicciones. A partir de este insumo, en la segunda parte se detallará la estrategia total propuesta para conformar el portafolio.

2.1 Estrategia propuesta para predecir el movimiento del precio de las acciones

2.1.1 Selección de acciones

Para la selección de activos que participarían en la conformación del portafolio, se tuvo en cuenta el índice bursátil COLCAP, puesto que éste es un índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la BVC (Banco de la República, s.f.). En este sentido, se seleccionaron las acciones que a hoy están activas y que han estado en la canasta de este índice en algún momento en el periodo 2013-2020, ya que ese es el periodo de análisis. La Tabla 2-1 muestra el listado de acciones que han ingresado a la canasta del COLCAP para el periodo de análisis.

Tabla 2-1: Acciones que han pertenecido al COLCAP en el periodo 2013-2020.

Acción	Estado	Decisión
PFBCOLOM	Activo	Seleccionada.
ECOPETROL	Activo	Seleccionada.
GRUPOSURA	Activo	Seleccionada.
NUTRESA	Activo	Seleccionada.
GRUPOARGOS	Activo	Seleccionada.
PFGRUPSURA	Activo	Seleccionada.
ISA	Activo	Seleccionada.

Tabla 2-1: (Continuación)

Acción	Estado	Decisión
PFAVAL	Activo	Seleccionada.
BCOLOMBIA	Activo	Seleccionada.
ÉXITO	Activo	Seleccionada.
CEMARGOS	Activo	Seleccionada.
PFDVVNDA	Activo	Seleccionada.
CORFICOLCF	Activo	Seleccionada.
EEB	Inactivo	Se descarta por inactividad actual de la acción.
BOGOTA	Activo	Seleccionada.
PFCEMARGOS	Activo	Seleccionada.
PFGRUPOARG	Activo	Seleccionada.
CELSIA	Activo	Seleccionada.
CLH	Activo	Seleccionada.
ISAGEN	Inactivo	Se descarta por inactividad actual de la acción.
PREC	Inactivo	Se descarta por inactividad actual de la acción.
GEB	Activo	Se descarta ya que es una acción muy nueva (inició en 2019).
INVERARGOS	Inactivo	Se descarta por inactividad actual de la acción.
PFAVH	Activo	Seleccionada.
CNEC	Activo	Seleccionada.
GRUPOAVAL	Activo	Seleccionada.
PROMIGAS	Activo	Seleccionada.
BVC	Activo	Se descarta ya que en 2018 sufrió un fraccionamiento inverso.
CONCRET	Activo	Seleccionada.
ETB	Activo	Seleccionada.
PMGC	Inactivo	Se descarta por inactividad actual de la acción.
PFAVTA	Inactivo	Se descarta por inactividad actual de la acción.
TERPEL	Activo	Seleccionada.
PFHELMBANK	Inactivo	Se descarta por inactividad actual de la acción.
FABRICATO	Activo	Seleccionada.
MINEROS	Activo	Seleccionada.

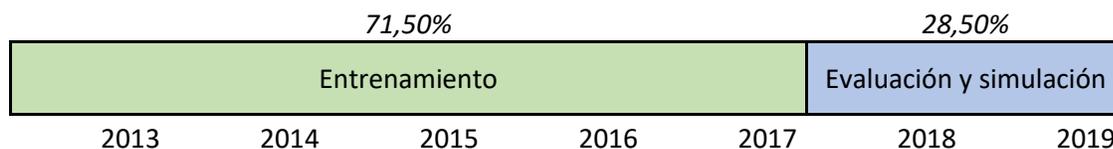
De acuerdo con lo anterior, se seleccionan 27 acciones para participar en la conformación del portafolio.

2.1.2 Selección de periodo de observación

El periodo a partir del cual se extraerán los datos de las acciones es 2013-2019. No se incluye el año 2020 debido a la pandemia del SARS-CoV-2 (COVID-19). El periodo de

análisis debe ser dividido en dos, una parte para entrenar el modelo y la siguiente parte es para testear el modelo y simular la estrategia de inversión. La Figura 2-1 presenta la división del periodo de análisis; el entrenamiento fue realizado entre el 1ro de enero de 2013 y el 31 de diciembre de 2017; el periodo de evaluación y simulación está comprendido entre el 1ro de enero de 2018 y el 31 de diciembre de 2019.

Figura 2-1: Periodo de análisis para entrenar y evaluar la predicción de la dirección del precio.



2.1.3 Definición del problema de pronóstico

Gracias a los acelerados avances tecnológicos, en los mercados financieros se ha vuelto común la construcción de herramientas de análisis técnico y cuantitativo que permitan desarrollar sistemas de apoyo a la toma de decisiones. Debido a lo anterior, las diversas técnicas en el campo del aprendizaje automático han tomado relevancia para el pronóstico de series de tiempo (Aloud, 2020).

Particularmente, en el mercado de valores, los títulos de renta variable son una alternativa de inversión que puede generar gran rentabilidad, pero es indiscutible que esto es directamente proporcional al nivel de riesgo o variabilidad que pueden presentar. En este sentido, para la inversión en este tipo de activos se ha vuelto relevante las proyecciones basadas en la información histórica. Es por esto que se han construido e implementado diversas técnicas de pronóstico; por un lado, se ha apostado a la predicción del precio del día siguiente; y por otro, se ha hecho uso, en gran medida, de la predicción direccional, es decir, del pronóstico del movimiento del precio (sube o baja).

En este trabajo, el problema de pronóstico se plantea como un problema de clasificación binaria (aprendizaje supervisado), lo que indica que se hará uso de la predicción direccional del precio (tendencia) de las acciones y no del valor continuo del precio del día siguiente. Se toma esta decisión dado que en la revisión de literatura realizada se encontró

que varias investigaciones han aplicado con éxito el pronóstico de tendencia del precio del día siguiente. Inclusive, se ha manifestado que las estrategias de negociación creadas a partir de pronósticos de la dirección del movimiento del precio y no de su valor, son más precisas, más efectivas y generan mayores rendimientos (Chen, Leung & Daouk, 2003; Huang, Nakamori & Wang, 2005; Choudhry & Garg, 2008). Esto, ha llevado a concluir que basta con que el inversor tenga en cuenta la dirección esperada del precio del día siguiente, puesto que le brindará las señales necesarias para saber tomar decisiones de compra o venta de una acción (Choudhry & Garg, 2008).

2.1.4 Definición de variables

2.1.4.1 Variables independientes

En la Tabla 2-2 se muestran las variables inicialmente planteadas para incluir en los modelos predictivos del movimiento del precio de las acciones. No obstante, más adelante, mediante una técnica de inclusión y exclusión de variables, se determinará cuáles son las variables significativas para el modelo en función de generar mayor rendimiento al inversionista.

Tabla 2-2: Variables seleccionadas para predecir dirección del precio (Rankia, 2020, BVC, s.f.-a).

Variable	Descripción
Precio cierre	Precio de cotización al cierre del día.
Precio mayor	Precio máximo de negociación del día.
Precio menor	Precio mínimo de negociación del día.
Volumen	Cantidad total de la actividad comercial del día.
Cantidad	Número de unidades del respectivo título o papel negociados en un día.
COLCAP	Es un indicador que refleja las variaciones de los precios de las 20 acciones más líquidas de la Bolsa de Valores de Colombia.
SPY (ETF S&P 500)	Este ETF es un fondo negociado en bolsa que cotiza en NYSE, el cual replica el índice S&P500 y por tanto las empresas más grandes y significativas de EE.UU. además, cuenta con una amplia trayectoria. Este fondo es el ETF más grande del mundo.

Dado que este es realmente un problema de series de tiempo, en el pronóstico del movimiento se deben tener en cuenta valores actuales y pasados de las variables presentadas en la Tabla 2-2. Un esquema representativo de las variables usadas para pronosticar el día $t + 1$ es presentado en la Tabla 2-3.

Tabla 2-3: Esquema representativo de las variables para pronosticar la dirección del precio en $t+1$.

P_t	P_{t-1}	P_{t-2}	...	P_{t-n}	V_t	V_{t-1}	...	V_{t-n}	Etiqueta
500	510	410		520	1000	1100		1050	1

Donde:

- P_t : precio de cierre del día actual
- V_t : Volumen negociado en el día actual

Según lo anterior, el reto también radicará en definir el retardo óptimo a utilizar para cada variable considerada.

2.1.4.2 Variable dependiente (respuesta)

Como se definió anteriormente, el problema a modelar es de clasificación y en este caso es binario, es decir, la variable respuesta está conformada por dos clases (0 y 1). Dado que el objetivo es predecir la tendencia del precio de las acciones del día siguiente, las clases se definieron de esta manera:

Clase 0: el precio de acción no subirá el día siguiente.

Clase 1: el precio de la acción subirá el día siguiente.

La predicción del movimiento del precio le permitirá al inversionista tomar decisiones de inversión, en donde se entrarán en posiciones largas (compra de acciones) cuando el precio de la acción suba el día siguiente (**clase 1**), y se tomarán posiciones cortas (venta de acciones) cuando la predicción del movimiento del precio indique que este no subirá (**clase 0**). La etiqueta fue calculada de la siguiente manera:

$$Etiqueta = Si(P_t < P_{t+1}, 1, 0) \quad (2.1)$$

Donde:

- P_t : precio de cierre del día actual.
- P_{t+1} : precio de cierre del día siguiente.

En la Tabla 2-4 se muestra un ejemplo de la manera cómo se definió la variable respuesta para entrenar los modelos.

Tabla 2-4: Ejemplo de la definición de la variable respuesta.

Acción	Fecha	Precio Cierre	Etiqueta (tendencia futura del precio)
ISA	2013-09-25	9.700	0 
ISA	2013-09-26	9.620	1 
ISA	2013-09-27	9.700	0 
ISA	2013-09-30	9.450	1 
ISA	2013-10-01	9.490	1 
ISA	2013-10-02	9.310	0 
ISA	2013-10-03	9.320	1 

2.1.5 Técnicas de pronóstico

Con el fin de encontrar la mejor técnica o combinación de éstas para la predicción del movimiento del precio de las acciones, se plantea el uso de varios algoritmos de clasificación (aprendizaje supervisado), los cuales serán implementados en Python. A continuación, se enuncian los métodos considerados:

- Regresión logística.
- Árbol de decisión.
- Bosque aleatorio.
- K vecinos más cercanos.
- Red neuronal recurrente LSTM (Long short-term memory).

Las cinco técnicas anteriores se aplicarán de manera individual a cada una de las 27 acciones seleccionadas, puesto que después en la estrategia del portafolio se definirá, con base en esta predicción, cuando comprar o vender una de dichas acciones. La idea entonces es definir qué modelo predice con mayor precisión la dirección del precio para

cada una de las acciones seleccionadas. En el siguiente capítulo se dará respuesta a este interrogante, estableciendo si se puede usar un mismo modelo para pronosticar los cambios de dirección de todas las acciones, o si, por el contrario, debería usarse un modelo diferente para cada acción.

2.1.6 Evaluación de los modelos

Con el fin de lograr la mayor precisión en la predicción de la dirección del precio de las acciones, los diferentes modelos de clasificación mencionados anteriormente serán evaluados con la métrica precisión equilibrada (también llamada precisión balanceada, del inglés *balanced accuracy*), la cual permite medir correctamente el desempeño del modelo independientemente del balanceo de las dos clases. Esta métrica estima el rendimiento del modelo en conjuntos de datos no equilibrados ya que se calcula como el promedio de la proporción de predicciones correctas de cada clase (Zhang, Li, Shahriar, Tao, Bhattacharya, & Qian, 2019). En este sentido, como se puede ver en la siguiente ecuación, la precisión equilibrada se basa en 2 métricas: sensibilidad y especificidad, lo cual es bastante beneficioso para el actual problema de clasificación ya que la predicción de ambas clases es igual de importante para poder tomar correctamente decisiones de compra y venta de una acción.

$$\text{Precisión equilibrada} = \frac{\frac{TP}{TP+FP} + \frac{TN}{TN+FN}}{2} = \frac{\text{Sensibilidad} + \text{Especificidad}}{2} \quad (2.2)$$

Donde:

TP, **FP**, **TN** y **FN** se obtienen de la matriz de confusión presentada en la Tabla 2-5.

Tabla 2-5: Matriz de confusión.

		Valores reales	
		1	0
Predicción	1	Verdaderos positivos (TP)	Falsos positivos (FP)
	0	Falsos Negativos (FN)	Verdaderos Negativos (TN)

2.2 Estrategia propuesta para la construcción del portafolio

Es importante resaltar que para lograr una mejor estrategia algorítmica de negociación no basta solo con construir un modelo predictivo, sino que es necesaria la intervención del inversionista o de especialistas en trading para definir la manera en que se utilizarán los resultados que entregan los modelos de pronóstico. Además, será necesaria la definición de otras reglas y otros parámetros que se dan en la negociación, como por ejemplo el mínimo valor de cada operación y el porcentaje del capital a invertir en cada movimiento. En este sentido, en esta sección se plantea la estrategia completa mediante la cual se construirá el portafolio de inversión.

2.2.1 Utilización del modelo predictivo de la dirección del precio para tomar decisiones de compra y venta de acciones

Los resultados de la aplicación de los modelos mencionados en la Sección 2.1.5 permitirán tener la predicción diaria del movimiento del precio de las acciones seleccionadas. Esto se convertirá en señales que le servirán al inversionista para definir en qué momento abrir o cerrar una posición de una determinada acción. En este orden de ideas, a partir de la predicción del modelo, se construye una variable S_t que señalará los momentos de compra y venta de una acción. Dicha variable se construye a partir de la siguiente ecuación:

$$S_t = \begin{cases} si & t = 1 & Pr_t \\ si & t \neq 1 & Pr_t - Pr_{t-1} \end{cases} \quad (2.3)$$

Donde:

- Pr_t : Predicción del movimiento del precio en t .
- S_t : Señal en t .

En la Tabla 2-6 se presenta un ejemplo de cómo se construye la señal. La columna Predicción corresponde al pronóstico del movimiento del precio Pr_t . La columna Señal hace referencia a S_t . La columna Decisión explica el resultado de la Señal. Cuando la Señal es **0** no se efectuarán operaciones financieras ya que el precio mantiene la misma

tendencia (alcista o bajista); cuando es **1** indica que la tendencia pasará de la baja al alza; y cuando finalmente la señal es **-1** indica que la tendencia del precio pasará del alza a la baja por lo cual se deberá vender la acción.

Tabla 2-6: Ejemplo de las decisiones de compra y venta de las acciones a partir de la predicción de la dirección del precio.

Predicción	Señal	Decisión
0	0	No efectuar operaciones.
0	0	No efectuar operaciones.
1	1	Compra.
1	0	No efectuar operaciones.
0	-1	Venta.
0	0	No efectuar operaciones.
1	1	Compra.
1	0	No efectuar operaciones.
1	0	No efectuar operaciones.
0	-1	Venta.

Con base en lo anterior, se plantean las siguientes reglas para tomar decisiones de inversión. Estas reglas aplican para todas las acciones, posteriormente se explicará en que orden se harán los movimientos.

Reglas:

1. Una predicción igual a 1 es una señal para la compra del activo. En este punto, el pronóstico refleja que el precio subirá el día siguiente lo que resulta rentable dado que para ese día ya habrá una ganancia.
2. La venta del activo se hará al encontrar el primer pronóstico que sea igual a 0 que se tenga después del día en que se efectuó la compra, puesto que el pronóstico dice que el precio no subirá el día siguiente por lo que es más rentable vender ya.
3. Cuando se compra un paquete de acciones de un activo, no se realizarán compras de más paquetes de acciones de este activo hasta que se venda el paquete que se

posee. Cuando se vendan las acciones de un activo, se venderá la totalidad de ellas.

4. Se podrán hacer múltiples operaciones de compra y venta durante todo el periodo de tiempo establecido, y desde que se tenga el dinero suficiente para invertir.
5. Se reinvierten las ganancias.
6. No se harán inyecciones adicionales al capital original con que se inicia la simulación.
7. La conformación del portafolio se actualizará todos los días en que opera la Bolsa de Valores de Colombia.

Es importante resaltar que, aunque la señal indique comprar una o varias acciones, no siempre se efectuará la compra ya que primero se deberá validar que haya disponible la suficiente cantidad de dinero para realizar la transacción. Esto se verá adelante con más detalle.

2.2.2 Priorización de compra de acciones

Luego de definir las señales de compra y venta de cada acción, se debe construir un dataset unificando todas estas señales con el fin conformar el portafolio diariamente. En este punto, es importante definir el orden en que se comprarán las acciones, puesto que para una misma fecha pueden existir varias señales de compra de diversas acciones; esto se justifica porque el capital es finito, y por lo tanto pueden existir momentos en que no se tenga el suficiente dinero para efectuar todas las operaciones de compra sugeridas en un día.

Dado lo anterior, se propone priorizar las operaciones de compra de las acciones según la participación que han tenido en el índice COLCAP en el periodo de análisis 2013-2019, donde una mayor participación significará mayor priorización. La Tabla 2-7 muestra dicha priorización, la cual fue construida tomando como referencia la suma del peso (participación) que ha tenido cada acción en el índice en el periodo especificado. Dado que el COLCAP toma las 20 acciones más líquidas, este peso se calcula con base en el número de acciones (en miles de millones) que tiene cada compañía en poder de los inversionistas (Bolsa de Valores de Colombia (BVC), 2011).

Tabla 2-7: Priorización de acciones para efectuar operaciones en el portafolio.

Activo	Suma de participación en COLCAP 20213-2019	Priorización
PFBCOLOM	378,8%	1
ECOPETROL	370,2%	2
GRUPOSURA	208,2%	3
NUTRESA	186,1%	4
GRUPOARGOS	182,4%	5
PFGRUPSURA	154,6%	6
ISA	138,0%	7
PFAVAL	122,2%	8
BCOLOMBIA	118,6%	9
EXITO	110,3%	10
CEMARGOS	104,1%	11
PFDVVNDA	82,3%	12
CORFICOLCF	81,8%	13
BOGOTA	71,0%	14
PFCEMARGOS	68,9%	15
PFGRUPOARG	66,2%	16
CELSIA	51,0%	17
CLH	48,7%	18
PFAVH	21,4%	19
CNEC	11,4%	20
GRUPOAVAL	7,7%	21
PROMIGAS	6,6%	22
CONCONCRET	3,6%	23
ETB	2,3%	24
TERPEL	0,8%	25
FABRICATO	0,4%	26
MINEROS	0,4%	27

En la Tabla 2-7 se observa que la primera acción a tener en cuenta a la hora de efectuar una operación de compra es PFBCOLOM (preferencial de Bancolombia) y la última es MINEROS debido a su escasa participación en el COLCAP.

2.2.3 Definición de parámetros en la negociación

Al momento de efectuar una operación financiera intervienen diferentes variables, como por ejemplo el precio al que se comprará o venderá una acción y el costo de transacción. Adicionalmente, de cara al inversionista también se asocian otros parámetros tales como el monto que invertirá o el porcentaje del capital disponible que destinará a tranzar en cada operación; su aversión al riesgo será representada por el porcentaje máximo que estará

dispuesto a perder (stop loss). A continuación, se muestra la manera en que se definen todos estos parámetros que intervienen en la negociación.

2.2.3.1 Precio de compra y venta de acciones

Se hará el supuesto que el precio tanto para comprar o vender acciones en un día determinado, será el precio de cierre de ese día.

2.2.3.2 Costo de transacción

Dado que solamente los intermediarios de valores pueden acceder a este mercado, el inversionista debe vincularse a una firma comisionista con el fin de poder realizar sus inversiones. Por lo anterior, estos intermediarios cobran al inversionista un costo asociado a la transacción; sin embargo, esta tarifa no está regulada, sino que está sujeta a la definición de cada firma (BVC, 2020). De esta manera, para efectos prácticos, se hará el supuesto que el costo de transacción es fijo e igual a 0,02% sobre el volumen total negociado en cada operación (compra o venta).

2.2.3.3 Capital inicial

Para efectos prácticos se fijará el capital inicial a invertir en 10.000.000 de pesos colombianos. Como se mencionó anteriormente, las ganancias a partir de este capital serán reinvertidas, pero no habrá inyección de capital adicional.

2.2.3.4 Porcentaje del capital disponible a invertir en cada operación

Debido a que el portafolio se va a conformar con las 27 acciones seleccionadas anteriormente, se hace necesario definir qué porcentaje del capital disponible será invertido en cada operación. Esta restricción impide invertir todo el capital disponible en una sola acción, lo cual es riesgoso ya que la rentabilidad depende exclusivamente de una compañía, por lo que la idea es diversificar el portafolio con la adquisición de diferentes acciones. De esta manera, se propone realizar un análisis de sensibilidad con el objetivo de encontrar ese porcentaje óptimo que indicará la proporción a invertir en cada operación, el cuál será determinado a partir del máximo ROI obtenido en la simulación del portafolio.

Sin embargo, para la selección de este porcentaje, se ve involucrada la aversión al riesgo del inversionista. En el siguiente capítulo, se detallará este análisis.

Adicional a este porcentaje, en la estrategia de conformación de portafolio se propone definir un valor mínimo a ser transado, ya que como es sabido, cada operación tiene unos costos de transacción asociados. Lo anterior implica que realizar muchas operaciones de poco monto podría no ser tan rentable. En este aspecto, con el fin de no invertir valores muy pequeños cuando se tiene poco capital disponible, se propone fijar el monto de 500.000 pesos como aquel valor mínimo aceptado para poder proceder con la compra de alguna acción. En caso tal de que el monto a invertir no sobrepase este valor, la operación no será llevada a cabo.

2.2.3.5 Pérdida máxima

La pérdida máxima (Stop Loss, en inglés) se ha definido como un mecanismo por medio del cual, el inversionista puede detener sus pérdidas (Rankia, 2017). En este orden de ideas, definir cuánto está dispuesto a perder el inversionista, depende básicamente de su aversión al riesgo; una persona conservadora tolerará solo pequeñas variaciones negativas en su portafolio, mientras que un inversionista con un perfil arriesgado está dispuesto aceptar grandes riesgos en búsqueda de obtener altas rentabilidades. En este sentido, para efectos prácticos, se definirá la pérdida máxima aceptada en un 3%. Lo anterior significa que, en la evaluación diaria del portafolio, se determinará la ganancia o pérdida acumulada que se tiene sobre cada una de las acciones, y en caso tal que la pérdida sobre la inversión sea superior al stop loss establecido, de inmediato se procederá con la venta total de la acción, aun cuando la señal para ese día no haya indicado la venta.

2.3 Conclusiones

En este capítulo se ha presentado la metodología que permitirá el desarrollo de un modelo de simulación de portafolio basado en una estrategia algorítmica de negociación. Dicha estrategia se crea a partir de la aplicación de modelos de machine learning que entregan predicciones sobre la tendencia del precio de las acciones, lo cual se traduce en señales

de compra y venta para la conformación de la cartera. En ese sentido, a lo largo de este capítulo se vio que la estrategia automática de trading involucra una parte netamente predictiva a cargo del entrenamiento de los modelos de IA, pero otra parte importante es el conocimiento que el inversionista o expertos de trading deben incorporar en la estrategia. Pues sí bien los modelos ayudan a automatizar la conformación del portafolio, también se hace necesaria la intervención humana en algunos puntos de la estrategia al definir otros factores como la pérdida máxima aceptada y el capital disponible a invertir, lo cual no lo establece automáticamente el algoritmo sino que está basado en la aversión al riesgo de cada individuo.

Finalmente, la metodología planteada en este capítulo pretende entregar al inversionista una herramienta valiosa para la toma de decisiones en cuanto a la conformación del portafolio de renta variable en el mercado financiero colombiano, la cual es apoyada en la potencia de diversos tipos de modelo de IA. Adicionalmente, con esta herramienta, se elimina el sesgo emocional que muchas veces tienen los inversionistas al momento de realizar operaciones financieras. En este orden de ideas, para evaluar la utilidad o el valor de esta metodología, en el siguiente capítulo se desarrollará el paso a paso de lo propuesto en este capítulo con el fin de mostrar la rentabilidad que tendría el inversionista al ejecutar esta metodología automática de inversión, y de esta manera se podrá determinar si es conveniente o no seguirla utilizando.

3. Resultados y Discusión

En esta sección se dará cumplimiento a los objetivos anteriormente planteados, donde bajo la metodología CRISP-DM se desarrollarán cada una de las etapas para la construcción de los modelos predictivos del movimiento del precio de las acciones. A partir de estos resultados, la segunda parte de este capítulo detalla la estrategia de trading automático llevada a cabo para definir la conformación diaria del portafolio, donde se incluye un modelo de simulación que evalúa diariamente la rentabilidad del portafolio. Adicionalmente, se propone comparar el desempeño de la estrategia de AT con respecto a la estrategia tradicional de conformación de portafolio (mínima varianza – Markowitz).

3.1 Definición del problema de negocio

En la literatura se ha señalado que el principal problema de investigación en la negociación del mercado de valores es la selección de una estrategia comercial para tomar decisiones en un instante de tiempo determinado. El tomar la decisión hoy o postergarla para otro día es vital dado que la rentabilidad depende del momento de compra o venta de un activo. En ese sentido, este desafío implica que la obtención de ganancias en este tipo de mercado tiene un alto nivel de incertidumbre, el cuál es inherente a la tendencia del precio de las acciones (Chakole, Kolhe, Mahapurush, Yadav & Kurhekar, 2021). Así, la predicción de la tendencia futura o del precio se convierte en un problema de investigación que deben resolver los inversionistas con el fin de saber cuándo tomar decisiones comerciales que aumenten su rentabilidad.

Dado lo anterior, el trading automático ha tomado cada vez más relevancia, y se ha posicionado como una herramienta que genera señales y ejecuta decisiones comerciales a través del uso de algoritmos informáticos, al tiempo que evita el sesgo emocional que muchas veces tienen los inversionistas al momento de comprar o vender una acción (Ma, Zhang, Liu, Ji & Gao, 2021). Adicionalmente, la tendencia del precio de las acciones cambia con tanta frecuencia, que un comerciante humano no tiene la capacidad de reaccionar instantáneamente para modificar sus decisiones de negociación.

En este contexto, surge la necesidad de crear estrategias de negociación automatizadas, donde los inversionistas o expertos del mercado sean quienes creen la lógica de negociación, pero finalmente el programa o máquina que se construye es capaz de aprender rápidamente del mercado, de las tendencias y de esta manera crea señales y ejecuta ordenes comerciales en un tiempo mucho inferior en comparación con un humano.

Debido a lo anterior y como se evidenció en la revisión de literatura, Estados Unidos le ha apostado en gran medida a este tipo de herramientas automatizadas, donde más del 70% del volumen de activos negociado en el país se da por AT. No obstante, en Colombia la situación es contraria y hay pocas evidencias de esfuerzos en este campo de investigación, por lo que resulta pertinente estudiar y probar el impacto que tiene AT en el mercado de valores de Colombia, a través de la construcción de modelos predictivos que generen señales de compra y venta de activos maximizando las ganancias para el inversionista.

3.2 Definición del problema de analítica

Como se mencionó en la Sección 2.1.3, en este trabajo el problema de pronóstico se plantea como un problema de clasificación binaria (aprendizaje supervisado), lo que indica que se hará uso de la predicción direccional del precio (tendencia) de las acciones y no del valor continuo del precio del día siguiente. Este problema es de corte supervisado ya que se cuenta con los datos de las series de tiempo de los activos financieros, por lo que es posible saber históricamente si el precio sube o baja al día siguiente. Las dos clases que involucran la variable respuesta están identificadas con el número 1 el cual indica que el precio subirá el día siguiente; y el número 0 que indica que el precio no subirá el día siguiente.

Lo anterior significa que los modelos que se construyan entregarán señales para generar diversas operaciones comerciales. Si para un día, el pronóstico estima que el precio subirá el día siguiente, esto se traducirá en una señal de compra del activo ya que la tendencia será alcista. Es importante resaltar que para cada acción podría tenerse un modelo diferente y, luego las señales que entreguen estos modelos se unirían para la conformación del portafolio. De esta manera, el contar con la predicción del movimiento del precio de las acciones se convierte en la base para formular estrategias comerciales automáticas que

le permitan al inversionista tomar decisiones de inversión. Adicionalmente, los modelos que se propongan tendrían la capacidad de aprender constantemente de los datos, por lo que la estrategia podría refinarse a medida que se evalué el modelo.

3.3 Información utilizada

La información del mercado de valores en Colombia la administra la BVC (Bolsa de Valores de Colombia, s.f.-b). Los datos de los activos de renta variable seleccionados se descargaron la página web de la BVC. Para cada acción, se extrajeron los siguientes datos históricos: precio de cierre, precio mayor, precio menor, volumen y cantidad negociada para cada día. En la extracción se obtuvo la información comprendida entre el primero de enero 2013 hasta el 31 de diciembre de 2019.

Adicionalmente a estos datos, también se extrajo de la BVC, el histórico del índice COLCAP para el mismo periodo de tiempo, ya que este índice se considerará más adelante como una variable a incluir en los modelos de pronóstico. Gracias a que la información en la BVC es pública, se construyó en Python un algoritmo de web scraping el cual se dirige a la página web de la Bolsa de Valores de Colombia y extrae los datos de las acciones de forma automática.

Por otra parte, dado que se considera el ETF SPY de Estados Unidos como una variable explicativa en los modelos de pronóstico, se procedió a descargar esta información a través de la librería *yfinance* disponible en Python. El periodo de tiempo considerado fue el mismo mencionado anteriormente.

Nota importante: las series de tiempo de las acciones seleccionadas son tomadas a modo de ejemplo con los datos disponibles de la Bolsa de Valores de Colombia, por lo que dicha información es netamente para uso práctico.

3.4 Preparación de datos e ingeniería de características

3.4.1 Entendimiento de los datos

En la Figura 3-1 se puede observar el precio diario de cierre de las acciones seleccionadas; en esta figura se pueden apreciar diferentes tendencias y rango de precios para dichas acciones, lo que sugiere que ellas tienen dinámicas muy diferentes. Esto implica que una única técnica de pronóstico no necesariamente funciona bien para todas las acciones; este aspecto será explorado en detalle más adelante. Para esto, se prueban diferentes técnicas para las series de las acciones y se evalúa su desempeño predictivo con el fin de seleccionar la mejor combinación de técnicas de IA. En las diferentes gráficas, la línea punteada separa el periodo de entrenamiento (lado izquierdo) vs. el periodo de test o backtesting (lado derecho).

Figura 3-1: Histórico precio diario de las acciones.





En la Figura 3-1 también es importante resaltar que hay pares de acciones que tienen un comportamiento casi idéntico, tal es el caso de GRUPOAVAL-PFAVAL, CEMARGOS-PFCEMARGOS, GRUPOARGOS-PFGRUPOARG, BCOLOMBIA-PFBCELOM y GRUPOSURA- PFRUPSURA, lo cual es debido a que algunas empresas, adicional a la emisión de acciones ordinarias, emiten también acciones preferenciales.

Al analizar las gráficas también se observan ciertos comportamientos que son comunes en la mayoría de las acciones. Por ejemplo, a finales del año 2018 hay una marcada tendencia bajista en el precio, lo cual corresponde al desplome de las bolsas en el mundo que ocurrió en dicho periodo de tiempo. Particularmente, para esa fecha (octubre 2018), el mercado estadounidense borró todas las ganancias acumuladas en ese año por la caída del Dow Jones en 600 puntos (2,41%). Los analistas manifestaron que esto fue debido en gran parte a la guerra comercial entre EE.UU. y China, y por el temor de los inversionistas al alza en el costo de los créditos por la Reserva Federal (FED, por sus siglas en inglés) (BBC, 2018a). En diciembre de ese mismo año, Wall Street sufrió la peor semana de los últimos 10 años, desde la crisis financiera de 2008 (BBC, 2018b).

Dado lo anterior, se puede inferir que los sucesos de las bolsas a nivel mundial, particularmente Estados Unidos, impactan el comportamiento de las acciones de la Bolsa de Valores de Colombia. Para corroborar esto, en la

Figura 3-2 se analiza el comportamiento histórico del índice COLCAP de Colombia vs. el ETF del índice S&P (SPY) de Estados Unidos, donde se puede ver que a partir del año 2016 estos dos índices tienen un comportamiento similar. Debido a esto, se planteó incluir el índice SPY como una de las variables para tener en cuenta en los modelos con el fin de determinar su impacto en el precio de las acciones.

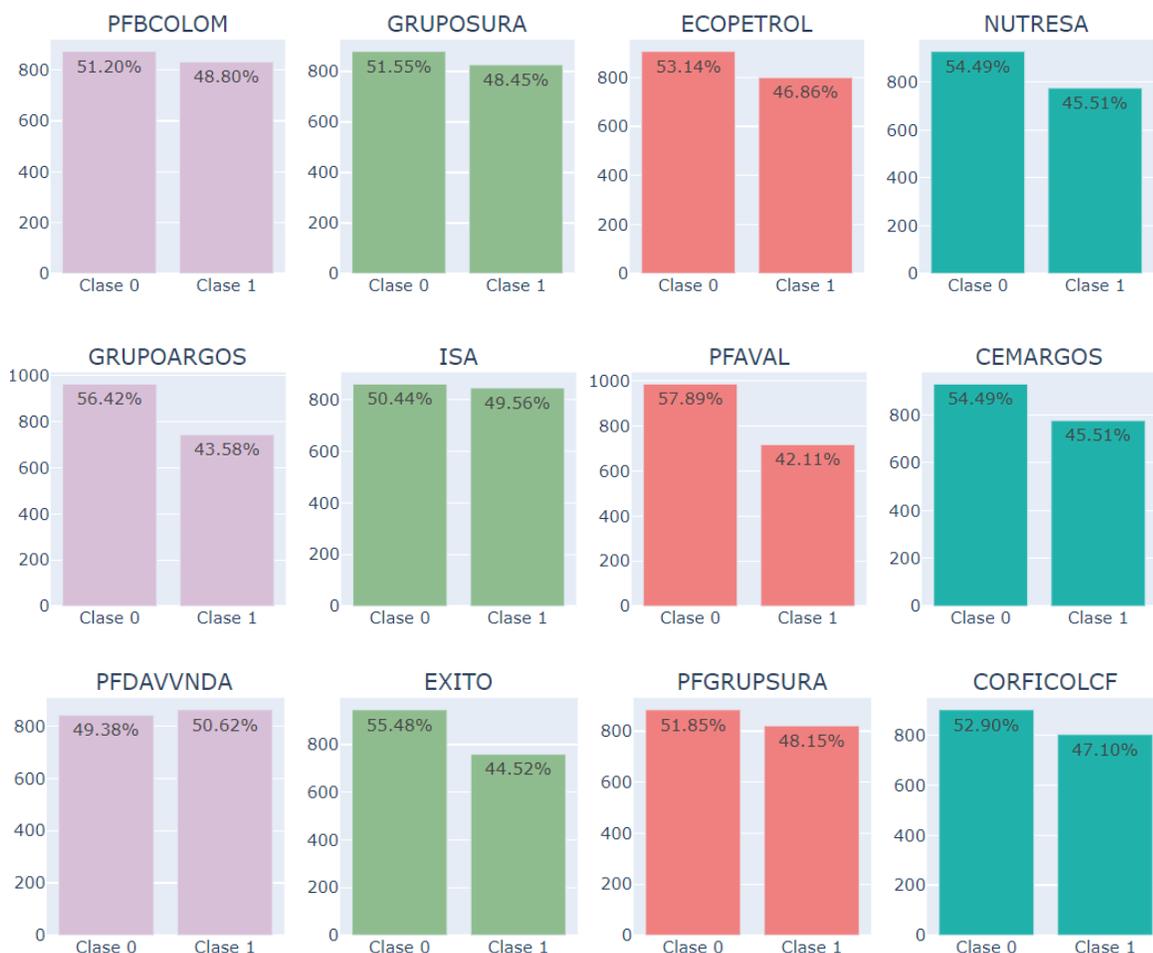
Figura 3-2: Índice Colcap vs. ETF del índice S&P (SPY).



Retomando la Figura 3-1, también se puede inferir que el número de tendencias bajistas y alcistas que presenta una misma serie varía considerablemente entre las acciones. Por

ejemplo, se puede observar que acciones como PROMIGAS, MINEROS y TERPEL presentan una tendencia predominantemente bajista con muy pocas recuperaciones. Para ver esto con más detalle, la Figura 3-3 muestra cual es la prevalencia de cada una de las dos clases en el problema de clasificación planteado (1: el precio sube el día siguiente y 0: el precio no sube), lo cual indica la predominancia de las tendencias (alcista o bajista). Como se puede ver, hay acciones como ISA, PFDVVNDA y BCOLOMBIA que son muy parejas en cuanto a las dos clases, sin embargo, hay otro tipo de acciones que tienen un desbalance muy marcado. Debido a esto, se decidió seleccionar la medida “balanced accuracy”, mencionada en la Sección 2.1.6, como aquella medida que evaluará el desempeño de los diferentes modelos para una adecuada selección, ya que esta métrica funciona muy bien tanto para datos balanceados como no balanceados.

Figura 3-3: Composición de las clases 0 y 1 en cada acción en el periodo 2013-2019.





3.4.2 Preparación de los datos

3.4.2.1 Variables y retardo

Como se mencionó en el Capítulo 2, inicialmente se plantean 7 variables explicativas del movimiento de cada precio. Cada activo tendrá su propio modelo, por lo que las variables de cada uno son las relacionadas a su comportamiento bursátil diario: precio de cierre, precio mayor, precio menor, volumen y cantidad negociada. Adicionalmente, se propone

probar para cada modelo otras dos variables que pueden afectar la tendencia del activo; la primera es el índice COLCAP que refleja la situación del mercado accionario colombiano, y la segunda es el ETF SPY de Estados Unidos, el cuál es el más grande del mundo, que replica el índice S&P 500 donde se encuentran las 500 empresas más grandes de EE. UU. Con respecto al máximo retardo a considerar para las variables, se plantea probar el comportamiento de cada modelo con 8 retardos diferentes (1,2,3,4,5,6,7 y 8).

A partir las 7 variables explicativas planteadas, se propone implementar una metodología para determinar la mejor combinación de variables y retardo en términos de la precisión del pronóstico de la tendencia. Esta metodología se explicará con detalle más adelante.

3.4.2.2 Normalización de las variables predictoras

Dado que la unidad de medida de las 7 variables es muy diferente, se propone realizar una normalización de los datos. Adicionalmente, ya que se pretende pronosticar la tendencia del precio, se ve involucrado el cambio porcentual de esta variable de un día a otro, por lo que para la normalización de las variables predictoras se decide también aplicar el cambio porcentual de cada una de las variables respecto al día anterior. De esta manera, un cambio porcentual positivo indicaría una tendencia alcista, y un cambio negativo reflejaría una tendencia bajista.

3.4.2.3 Definición de variable respuesta

Como se mencionó anteriormente, el problema de analítica del presente trabajo consiste en predecir el movimiento del precio para generar señales comerciales; en tal sentido, la variable respuesta involucra el cambio que tiene el precio del activo financiero de un día a otro. Adicionalmente, el supuesto realizado es que el precio de cierre es el precio para comprar o vender acciones; consecuentemente, el precio de cierre también se usa para construir la variable respuesta.

En este contexto, dado que no se pretende modelar la serie de tiempo de las acciones, sino que el análisis se convierte en un problema de clasificación, la variable respuesta debe ser de tipo categórica, y en este caso dicotómica (dos clases): 1 sube el precio el día siguiente y 0: el precio no sube el día siguiente. En este orden de ideas, y teniendo en

cuenta que el modelo es predictivo, es necesario saber que va a ocurrir el día siguiente con la tendencia del precio. Por tal motivo, para construir la variable respuesta con el fin de entrenar los modelos, se tendrá en cuenta lo que sucede con el precio el día siguiente: sube o no sube. Recapitulando lo mencionado en la metodología, en la Sección 2.1.4 se expuso la ecuación a partir de la cual se origina la etiqueta o la variable respuesta, tal como se muestra a continuación:

$$\text{Variable respuesta} = Si(P_t < P_{t+1}, 1, 0) \tag{3.1}$$

Donde:

- P_t : precio de cierre del día actual
- P_{t+1} : precio de cierre del día siguiente

La Tabla 3-1 muestra un ejemplo de la lógica de la variable respuesta. Se puede observar claramente que cuando el precio de la acción sube el día siguiente, la variable respuesta es 1 y en caso de que el precio sea igual o inferior al día siguiente, el valor será 0.

Tabla 3-1: Ejemplo construcción variable respuesta.

Fecha	Precio de Cierre	Variable respuesta	Justificación
2/01/2013	9.600	0	El precio no sube el día siguiente
3/01/2013	9.600	1	El precio sube el día siguiente
4/01/2013	9.670	0	El precio no sube el día siguiente
8/01/2013	9.330	1	El precio sube el día siguiente
9/01/2013	9.570	0	El precio no sube el día siguiente
10/01/2013	9.450	1	El precio sube el día siguiente
11/01/2013	9.590	0	El precio no sube el día siguiente
14/01/2013	9.500		

3.5 Metaheurística propuesta para usar los modelos de pronóstico en la conformación del portafolio

A continuación, se presenta la metaheurística desarrollada para la construcción de los modelos de pronóstico de la dirección del precio. Esta metaheurística usa como función objetivo la maximización del rendimiento obtenido por el inversionista. En ella se considera la combinación entre tipos de modelo, el porcentaje a invertir en cada movimiento, las variables utilizadas y sus retardos. En cada paso de la metaheurística propuesta se probarán combinaciones de diferentes modelos y variables con su retardo. La metaheurística entregará la mejor combinación de tipo de modelo, el porcentaje a invertir en cada movimiento, variables y retardos que se obtengan durante la ejecución de la metaheurística. Esta metaheurística es necesaria dado que computacionalmente no es viable realizar la búsqueda exhaustiva de la mejor combinación de las variables mencionadas.

Es importante resaltar que, para medir el desempeño de las combinaciones, no se evaluará la precisión alcanzada por los modelos, sino la ganancia obtenida al final del periodo de simulación; los modelos son utilizados dentro de la estrategia de negociación para generar las señales de compra y venta de los activos. Se decide evaluar de esta manera ya que lo que realmente interesa en este problema de negocio es la rentabilidad que se genera para el inversionista al aplicar reglas de negociación basadas en los modelos.

Nota: para efectos prácticos, para la simulación de portafolio de los pasos 2, 3 y 4 se fija en 40% el porcentaje del capital disponible a invertir en cada operación.

Los pasos de la metaheurística propuesta son los siguientes:

1. **Selección de características y retardo:** se elegirá una acción que sea representativa del comportamiento del mercado; en este caso se busca que sea similar a la tendencia que presenta el COLCAP. Para esta acción se probarán todas las combinaciones de variables explicativas (tales como el precio de cierre o el volumen) y de sus retardos posibles, con el fin de encontrar el conjunto de predictores más adecuado. En este paso se seleccionará un único retardo para

todas las variables seleccionadas; es claro que cada variable podría tener su propio retardo óptimo, pero en este caso se busca obtener un modelo base más que un modelo óptimo. La combinación óptima de variables explicativas y retardo para la acción seleccionada será usada para construir los modelos base de las demás acciones. Es decir, se usarán las mismas variables explicativas y el retardo seleccionado para las acciones restantes del portafolio. En este paso, se considera únicamente el modelo de regresión logística para realizar la predicción del movimiento de cada acción. Como criterio de selección de la mejor combinación de variables y retardo se utilizará la precisión balanceada, la cual fue definida en la Sección 2.1.6 de la metodología.

- 2. Probar si un mismo tipo de modelo en cada una de las acciones entrega señales de compra y venta que generan rentabilidad para el inversionista en la simulación del portafolio:** a partir de la combinación de variables y el retardo seleccionados en el punto anterior, se procederá a construir modelos de pronóstico para todas las acciones. En este paso se considera que siempre se usa el mismo tipo de modelo para el pronóstico de los activos, pero se busca el mejor tipo de modelo entre los posibles. Se considerarán los siguientes tipos de modelo para realizar la predicción:

- Regresión logística.
- Árboles de decisión.
- Bosques aleatorios.
- Vecinos más cercanos.
- Red neuronal recurrente.

Luego de entrenar los modelos, se procede a realizar la simulación del portafolio; los modelos estimados en este paso entregan las señales de compra y venta de los activos; en este punto es necesario enfatizar que en cada simulación se usa un único tipo de modelo, lo que implica cinco simulaciones. A partir del rendimiento obtenido en cada simulación, se selecciona el mejor modelo. Cada simulación tiene como objetivo determinar la ganancia que lograría el inversionista con cada una de las versiones según la técnica seleccionada. En otras palabras, como la función objetivo de la metaheurística es la maximización del rendimiento obtenido por el inversionista, se selecciona el tipo de modelo que logre mayor ganancia en la simulación. Este paso tiene como objetivo determinar si es posible utilizar un único

tipo de modelo para pronosticar el movimiento de los precios sin sacrificar la rentabilidad del inversionista. El resultado obtenido será comparado con el siguiente paso.

- 3. Probar si un tipo diferente de modelo para cada acción genera más rentabilidad para el inversionista que el utilizar un mismo tipo de modelo para todas las acciones:** en este punto se propone encontrar el modelo con mejor desempeño predictivo para cada acción. En este paso, se selecciona para cada acción el tipo de modelo que tenga la mayor precisión balanceada. Posteriormente se realiza la simulación del portafolio y se evalúa si la ganancia generada es superior o inferior al tipo de modelo seleccionado en el paso anterior. Dado que cada acción puede ser pronosticada por un tipo de modelo diferente, es posible establecer si esta política tiene resultados significativamente diferentes respecto a usar el mismo tipo de modelo para todas las acciones. En este paso se mantiene la combinación de variables y retardo seleccionados en el Paso 1.
- 4. Evaluar diferente número de acciones a incluir en la conformación del portafolio:** en las simulaciones anteriores, la conformación del portafolio se realiza con las 27 acciones seleccionadas; sin embargo, en este punto se decide evaluar y comparar la rentabilidad al conformar portafolios desde 1 hasta 27 acciones, con el fin de determinar un número apropiado de activos a considerar para la conformación del portafolio que le entreguen al inversionista una mayor rentabilidad
- 5. Variar el porcentaje del capital disponible a invertir en cada operación:** de acuerdo con lo mencionado en la Sección 2.2.3, en este paso varía el porcentaje a invertir en cada operación (de 10% a 100% en pasos de 10%) con el fin de encontrar un óptimo local que le favorezca al inversionista.

Finalmente, la metaheurística descrita permitirá encontrar un óptimo local en función de la maximización del rendimiento por el inversionista, de acuerdo con la combinación de variables, retardo y porcentaje del capital disponible a invertir en cada operación. Los tipos de modelos seleccionados entregarán señales de compra y venta de los activos que se utilizarán en la simulación del portafolio; esto se hará con el fin de estimar el retorno sobre la inversión generada y de esta manera poder determinar si la estrategia de trading algorítmico entrega mejores resultados en comparación con otras estrategias tradicionales, como por ejemplo la de mínima varianza.

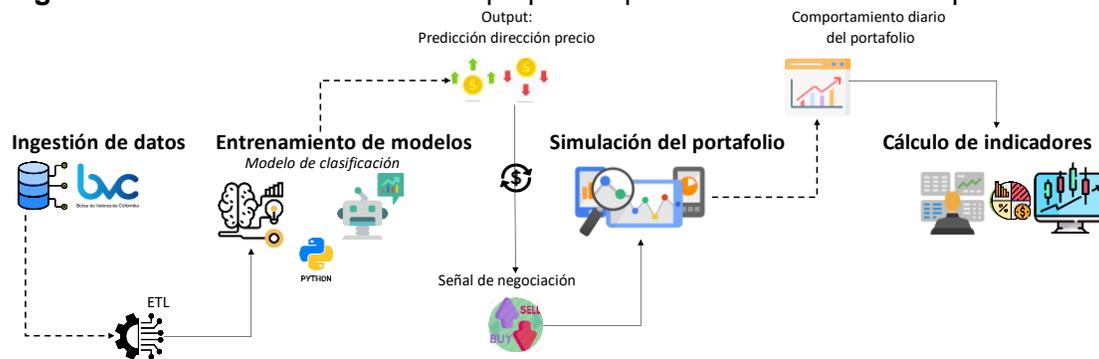
3.6 Resultados obtenidos y discusión

En esta sección se presenta los resultados numéricos de la metaheurística descrita en la Sección 3.5.

3.6.1 Modelo de simulación

Como se mencionó en la metodología, el modelo de simulación propuesto para la conformación del portafolio abarca diversos pasos, los cuales se detallan en la Figura 3-4. En dicha figura se observa que la simulación consta de 4 fases; ingestión de datos, entrenamiento de modelos, simulación del portafolio y finalmente cálculo de indicadores. Como se podrá ver en las siguientes secciones, los resultados son computados con distintas versiones de este modelo, de acuerdo a variaciones que realizan en los pasos 2 y 3 (entrenamiento de modelos y simulación de portafolio). El objetivo de realizar diversas simulaciones de este modelo es determinar la estrategia algorítmica de negociación que genera más rentabilidad para el inversionista.

Figura 3-4: Modelo de simulación propuesto para la conformación del portafolio.



3.6.2 Selección de características y retardo

En esta sección se desarrolla el Paso 1 de la metaheurística. Como se mencionó anteriormente, se consideran las siguientes siete variables explicativas para la construcción de los modelos:

- Precio cierre.
- Precio mayor.
- Precio menor.
- Volumen.
- Cantidad.
- COLCAP.

- SPY.

y los retardos del uno al ocho como posibles. Los modelos de pronóstico considerados son los siguientes:

- Regresión logística.
- Árboles de decisión.
- Bosques aleatorios.
- Vecinos más cercanos.
- Red neuronal recurrente.

Lo anterior significa que, si se quisiera determinar la mejor combinación de tipo de modelo, variables y retardo por cada acción, se tendrían que realizar 137.160 modelos diferentes.

$$\text{Modelos} = 27 \text{ acciones} \times 5 \text{ técnicas} \times 127 \text{ combinaciones variables} \times 8 \text{ retardos} = 137.160 \text{ modelos diferentes} \quad (3.2)$$

Las 127 combinaciones de variables en la ecuación anterior se obtienen al considerar todas las combinaciones posibles de variables explicativas; estas combinaciones abarcan todos los conjuntos que contienen desde una hasta siete variables explicativas. Para efectos de esta tesis, resulta inviable considerar la evaluación de 137.160 modelos, ya que, si la ejecución de un modelo toma un tiempo promedio de 2 segundos, se requerirían 76,2 horas para ejecutar todos los modelos. Por esta razón se decidió implementar una estrategia heurística menos costosa que permita encontrar un óptimo local para la selección del retardo y la combinación de variables, pero que tenga un desempeño apropiado para el inversionista; esto implica, que se busca obtener un óptimo local cercano al óptimo global que se obtendría al realizar una búsqueda exhaustiva sobre los 137.160 modelos.

Dado lo anterior, como se mencionó en la descripción de la metaheurística, lo primero que se debe realizar es la selección de una acción que sea representativa del comportamiento del mercado. Al realizar el análisis descrito en los párrafos anteriores, se encontró que ECOPETROL es la acción que durante el periodo de análisis ha estado mayor número de veces en el primer puesto según su participación en el COLCAP. Sin embargo, PFBCOLOM (Preferencial de Bancolombia) es la acción que tiene mayor participación

acumulada, es decir, al sumar el peso o la participación que ha tenido esta acción en cada canasta del COLCAP, se ubica como el activo con mayor suma de participación, tal como se observó en la Tabla 2-7. En la Figura 3-5 se puede apreciar el comportamiento que ha tenido el índice COLCAP vs las acciones de ECOPETORL y PFBCOLOM, donde se nota que esta última acción tiene una dinámica más similar al índice en comparación con ECOPETROL. Debido a lo anterior, se decide probar las combinaciones de variables y retardos en estas dos acciones, utilizando la regresión logística como modelo base; esto se realiza con el fin de seleccionar aquella combinación que tenga la mayor precisión en el pronóstico del movimiento del precio para ambas acciones.

Figura 3-5: Comportamiento COLCAP vs ECOPETROL y PFBCOLOM.



De acuerdo con el contexto anterior, se ejecutan 1.016 modelos diferentes de regresión logística para cada acción (ECOPETROL y PFBCOLOM), que resultan a partir de las 127 posibles combinaciones de variables y los 8 retardos planteados. La combinación óptima de variables explicativas y retardo para la acción seleccionada será usada para especificar los modelos base de las demás acciones; es decir, estos modelos usan las mismas variables y el mismo retardo, pero difieren en los parámetros particulares estimados para

cada acción. Para la selección óptima de las variables y retardo se propone evaluar los modelos con la medida “precisión balanceada”. La ejecución de los 2.032 modelos (= 1.016 modelos × 2 acciones) tardó un tiempo de 113 minutos con 21 segundos. En la Tabla 3-2 se presentan las mejores 10 combinaciones obtenidas por cada acción.

Por una parte, se tiene que, para ECOPETROL, la combinación que logra la mayor precisión en la predicción del movimiento del precio está conformada por las variables precio más alto, precio de cierre y el SPY, y un retardo de 6. Por otro lado, para PFBCOLOM, la mejor combinación es un retardo de 1 en las variables volumen y SPY. Sin embargo, puede observarse que entre las mejores 10 combinaciones de cada acción no hay una diferencia significativa en el desempeño predictivo, pues la precisión balanceada varía en un rango muy pequeño. Por lo anterior, a partir de estos resultados se procede a seleccionar las variables precio más alto, precio de cierre y el SPY con un retardo de 6 como la combinación que puede lograr un mejor desempeño en general para todas demás acciones, ya que dicha combinación es la mejor para predecir el movimiento del precio de ECOPETROL y se encuentre en el tercer lugar para PFBCOLOM.

Tabla 3-2: Resultados de combinaciones de variables y retardo para la predicción de la dirección del precio de ECOPETROL y PFBCOLOM.

ECOPETROL

Retardo	Variabes	Balanced Accuracy
6	['High', 'Close', 'SPY']	54,8%
5	['High', 'Low']	54,7%
6	['High', 'Low']	54,7%
2	['Trades', 'Volume']	54,7%
8	['High', 'Low', 'COLCAP']	54,7%
2	['Trades', 'Volume', 'COLCAP']	54,7%
2	['High', 'Trades', 'Volume', 'COLCAP']	53,4%
5	['High', 'Close', 'SPY']	53,2%
7	['High', 'Low', 'COLCAP']	53,0%
4	['High', 'Close', 'SPY']	53,0%

PFBCOLOM

Retardo	Variabes	Balanced Accuracy
1	['Volume', 'SPY']	56,3%
1	['SPY']	55,4%
7	['High', 'Close', 'SPY']	54,5%
2	['High', 'Close', 'SPY']	54,5%
7	['High', 'Close', 'Trades', 'SPY']	54,3%
7	['High', 'Close', 'Volume', 'SPY']	54,3%
2	['Low', 'Close', 'Trades', 'SPY']	54,2%
5	['High', 'SPY']	54,2%
3	['COLCAP', 'SPY']	54,1%
8	['High', 'Close', 'SPY']	54,1%

3.6.3 Estimación de los parámetros del modelo para todas las acciones del portafolio

3.6.3.1 Pronóstico usando el mismo tipo de modelo para todo el portafolio

En esta sección se desarrolla el Paso 2 de la metaheurística, en el cual, a partir de la combinación de variables y retardo seleccionados anteriormente (precio más alto, precio de cierre y el SPY con un retardo de 6) se procede a determinar cuál es el mejor modelo cuando este único tipo de modelo es usado para pronosticar todo el portafolio. En este paso se iniciará con la estimación del modelo de regresión logística, el cuál es usado como modelo base. En este punto surge la duda si la regresión logística es el mejor modelo para pronosticar el movimiento del precio de todas las acciones, o, en otras palabras, lo que se desea determinar es cuál es el modelo que da mejor rendimiento para el inversionista cuando éste se usa para pronosticar la dirección del precio de todas las acciones. Es decir, se usa el mismo tipo de modelo para pronosticar todos los movimientos, y este difiere de una acción a otra únicamente en sus parámetros. Para esto, se harán cinco simulaciones del portafolio, una por cada tipo de modelo (incluyendo la regresión logística) con el fin de evaluar la efectividad que tienen estas estrategias de trading algorítmico. Al ejecutar este paso, se realizaron 135 entrenamientos que corresponden a cinco tipos de modelos diferentes parametrizados para 27 acciones. El tiempo de ejecución fue 21 minutos con 16 segundos.

En la Tabla 3-3 se presenta la ganancia obtenida durante el periodo de simulación al usar un mismo tipo de modelo; en ella se puede observar la ganancia parcial y neta (después de descontar el costo de transacción) que tendría el inversionista el usar un mismo tipo de modelo para predecir el movimiento del precio de todas las acciones. Se puede ver que los bosques aleatorios son el tipo de modelo que logra el mayor rendimiento para el inversionista obteniendo una rentabilidad del 16% sobre la inversión; seguidamente, se encuentra la regresión logística y vecinos más cercanos con una rentabilidad de 15% y 8% respectivamente. Por otra parte, al usar los árboles de decisión o las redes neuronales recurrentes como único tipo de modelo en todas las acciones, el inversionista obtiene pérdidas (-1% y -7% respectivamente), por lo cual en esta simulación no es rentable usar estos tipos de modelo para todos los activos. Es importante recordar que, para efectos de la simulación, la inversión se estableció en 10.000.000 de pesos (ver Sección 2.2.3). Para

efectos prácticos, para la simulación de portafolio se fija en 40% el porcentaje del capital disponible a invertir en cada operación.

Tabla 3-3: Ganancia al usar un mismo tipo de modelo en cada acción.

Tipo modelo	Ganancia	Costo transacción	Ganancia Neta
Regresión logística	2.233.127	768.337	1.464.789
Red Neuronal Recurrente -LSTM	- 780.095	10.625	- 790.720
Vecinos más cercanos	1.741.577	908.256	833.321
Árboles de decisión	584.062	718.336	- 134.275
Bosques aleatorios	2.214.635	650.900	1.563.735

Por otra parte, en la Tabla 3-4 se presenta la precisión balanceada por tipo de modelo y por acción. Es necesario enfatizar que esta tabla se refiere a la precisión del pronóstico de cambio de dirección de precio del activo, y que no necesariamente el modelo más preciso en el cambio de dirección es el que genera más rendimientos. Como se puede observar, no hay un tipo de modelo que logre mayor precisión balanceada para todas las acciones, sino que por el contrario la tabla muestra que el tipo de modelo que tiene mayor precisión puede variar de una acción a otra. Tampoco hay un modelo que sea consistentemente inferior para todas las acciones.

Tabla 3-4: Precisión balanceada por tipo de modelo y por acción.

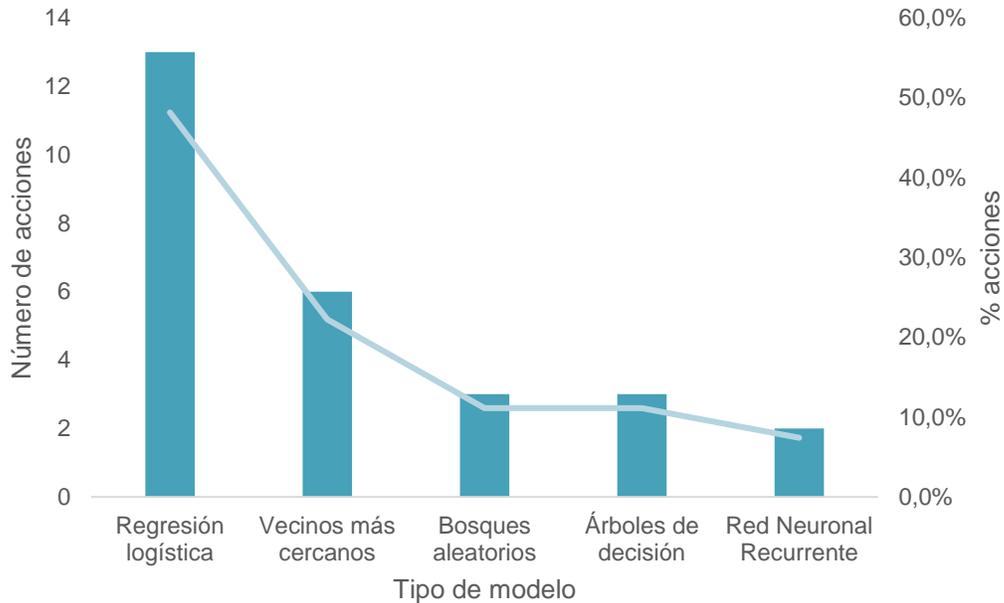
Acción	Vecinos más cercanos	Red Neuronal Recurrente	Bosques aleatorios	Regresión logística	Árboles de decisión
BCOLOMBIA	47,7%	50,0%	48,7%	52,3%	49,0%
BOGOTA	53,8%	50,0%	53,5%	58,2%	53,7%
CELSIA	50,5%	50,0%	50,0%	48,2%	50,0%
CEMARGOS	53,0%	50,0%	48,5%	52,2%	47,5%
CLH	51,8%	49,1%	50,0%	53,8%	50,5%
CNEC	48,2%	50,0%	48,8%	48,5%	49,9%
CONCRET	49,6%	50,0%	50,0%	54,7%	50,6%
CORFICOLCF	49,5%	50,0%	48,6%	51,2%	49,4%
ECOPETROL	53,5%	50,0%	50,2%	54,8%	52,0%
ETB	53,1%	50,0%	49,3%	48,6%	53,6%

Tabla 3-4: (Continuación)

Acción	Vecinos más cercanos	Red Neuronal Recurrente	Bosques aleatorios	Regresión logística	Árboles de decisión
EXITO	51,0%	50,0%	50,2%	55,0%	50,2%
FABRICATO	51,2%	50,0%	50,0%	55,4%	50,8%
GRUPOARGOS	49,8%	50,0%	50,6%	49,7%	49,8%
GRUPOAVAL	50,0%	50,0%	50,0%	55,5%	52,2%
GRUPOSURA	51,4%	50,0%	55,2%	54,4%	51,3%
ISA	52,3%	50,0%	50,7%	52,0%	51,4%
MINEROS	52,4%	50,0%	50,0%	54,1%	53,6%
NUTRESA	48,7%	50,0%	51,9%	45,4%	53,0%
PFAVAL	50,5%	50,0%	49,0%	56,3%	49,5%
PFAVH	46,1%	50,0%	49,6%	48,4%	50,0%
PFBCOLOM	46,4%	50,0%	49,9%	52,9%	51,0%
PFCEMARGOS	51,9%	50,0%	50,4%	50,7%	48,1%
PFDVVNDA	49,9%	50,0%	52,4%	51,4%	51,4%
PFGRUPOARG	54,4%	50,0%	51,7%	51,0%	49,1%
PFGRUPSURA	51,0%	50,0%	52,2%	53,8%	51,5%
PROMIGAS	51,5%	50,0%	50,0%	50,0%	51,6%
TERPEL	53,9%	50,0%	50,0%	50,8%	50,0%

De acuerdo con lo anterior, en la Figura 3-6 se presenta un gráfico de barras en el cual se cuenta la frecuencia en que cada modelo fue el más preciso en el pronóstico de la dirección del precio para cada acción. De la figura se observa que el modelo de regresión logística fue el más preciso para 13 acciones (48%), vecinos más cercanos para 6 (22%), árboles de decisión para 3 (11%), bosques aleatorios para 3 (11%) y, finalmente, redes neuronales recurrentes para 2 (8%). Esto puede explicar la razón por la que usar un mismo tipo de modelo para cada acción puede no generar rentabilidad, tal como el caso de las redes neuronales recurrentes y los árboles de decisión como se vio en la Tabla 3-3.

Figura 3-6: Tipos de modelo con el número de acciones que logra el mejor desempeño predictivo.



3.6.3.2 Pronóstico usando diferente tipo de modelo para cada acción

Según lo analizado en la sección anterior, se pudo determinar que un mismo tipo de modelo no tiene el mejor desempeño predictivo para todas las acciones, por lo cual usar un solo tipo de modelo podría no ser rentable para el inversionista. De esta manera, en esta sección se desarrolla el Paso 3 de la metaheurística donde se realiza la simulación del portafolio considerando diferentes tipos de modelos para las acciones, en donde la selección del tipo de modelo para cada acción se determina con base en el más alto desempeño predictivo del movimiento del precio. En la Tabla 3-5 se indica el tipo de modelo que será usado para el pronóstico del movimiento del precio de cada acción.

Tabla 3-5: Mejor tipo de modelo por acción para predecir la dirección del precio.

Acción	Vecinos más cercanos	Red Neuronal Recurrente	Bosques aleatorios	Regresión logística	Árboles de decisión
BCOLOMBIA				✓	
BOGOTA				✓	
CELSIA	✓				
CEMARGOS	✓				
CLH				✓	
CNEC		✓			
CONCONCRET				✓	
CORFICOLCF				✓	
ECOPETROL				✓	
ETB					✓
EXITO				✓	
FABRICATO				✓	
GRUPOARGOS			✓		
GRUPOAVAL				✓	
GRUPOSURA			✓		
ISA	✓				
MINEROS				✓	
NUTRESA					✓
PFAVAL				✓	
PFAVH		✓			
PFBCOLOM				✓	
PFCEMARGOS	✓				
PFDVVNDA			✓		
PFGRUPOARG	✓				
PFGRUPSURA				✓	
PROMIGAS					✓
TERPEL	✓				

En este orden de ideas, se procedió a hacer uso de las señales de compra y venta predichas para las acciones, según los resultados de los modelos señalados en la Tabla 3-5. Esto se realiza con el fin de hacer la simulación del portafolio y determinar la ganancia

o pérdida para el inversionista y poder comparar con los resultados de la sección anterior donde se utilizó un único tipo de modelo para todas las acciones. Es importante resaltar que estos modelos fueron entrenados con base en la combinación de variables y retardo seleccionados anteriormente (precio más alto, precio de cierre y el SPY con un retardo de 6).

Dado que los modelos ya habían sido entrenados en la sección anterior, en este caso solo se realizó la simulación del portafolio a partir de sus pronósticos; la simulación requirió un tiempo de ejecución de 3 minutos con 6 segundos. En la Tabla 3-6 se presentan los resultados obtenidos en la simulación. En la tabla se puede observar la ganancia parcial y neta (después de descontar el costo de transacción) que tendría el inversionista el usar el mejor modelo para cada acción con base en el desempeño predictivo. Se puede apreciar que la rentabilidad obtenida es del 36%, muy superior a la conseguida al utilizar un único modelo, donde la mejor fue del 16% (bosques aleatorios, ver Tabla 3-3). De esta manera, se puede inferir que resulta más rentable para el inversionista encontrar el mejor modelo para cada acción en lugar de usar un único tipo para todos los activos; esto en parte se debe al comportamiento particular de cada serie ya que la tendencia varía considerablemente entre las acciones tal como se observó en la Figura 3-1.

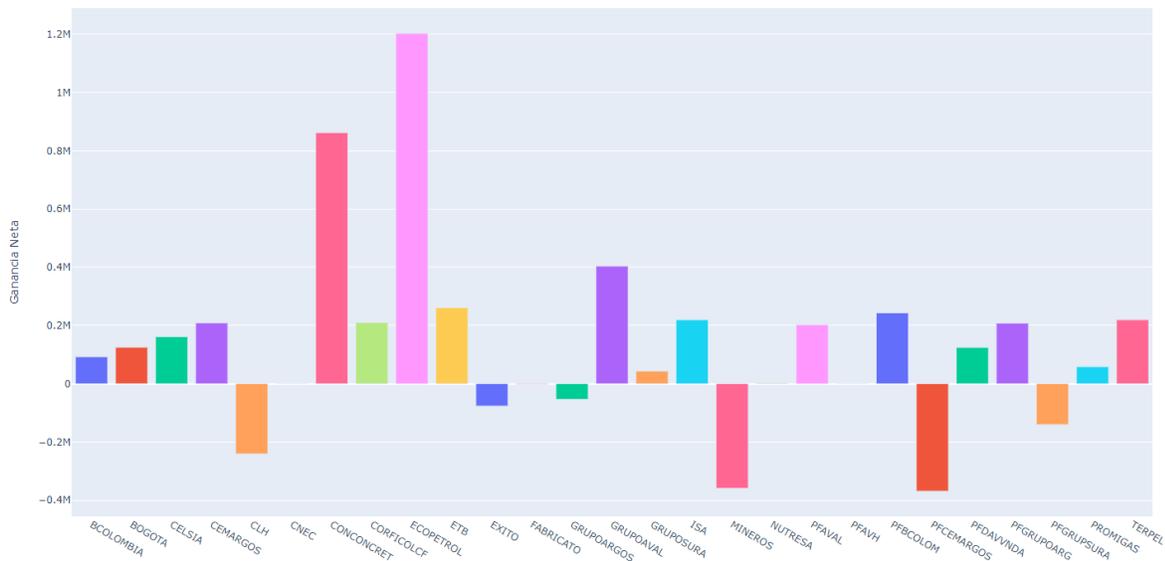
Tabla 3-6: Simulación portafolio utilizando diferentes modelos

Inversión	Ganancia	Costo transacción	Ganancia Neta	Rentabilidad
10.000.000	4.473.002	867.362	3.605.640	36%

En la Figura 3-7 se puede apreciar la evolución del valor del portafolio a lo largo del periodo de simulación, donde se puede ver que durante el primer año (2018), el valor del portafolio no presentó cambios muy significativos, aunque se observa que hubo dos periodos de pérdida. Caso contrario pasó durante el segundo año (2019), donde la recuperación del valor del portafolio fue evidente y hubo una notoria alza en el valor, manteniendo una tendencia alcista y alcanzando una ganancia de 3,6 millones de pesos aproximadamente.

Figura 3-7: Comportamiento valor portafolio en el tiempo

Es importante resaltar que el valor del portafolio anterior está conformado por todos los activos, por lo que resulta pertinente saber qué acciones son las que están entregando más rentabilidad al portafolio. Para esto, en la Figura 3-8 se presenta la ganancia total obtenida por acción, donde ECOPETROL resalta como el activo que genera mayor rentabilidad, superando los 1,2 millones de pesos de utilidad, lo cual representa aproximadamente el 33% del total de ganancias obtenidas en el portafolio de las 27 acciones. En segundo lugar, se encuentra CONCRET, con una ganancia cercana a los 900.000 de pesos, representando alrededor del 25% del total de la rentabilidad. En este sentido, se tiene que las dos acciones anteriores establecen el 58% del total de las ganancias obtenidas en la simulación del portafolio. En contraste, CLH, EXITO, FABRICATO, GRUPOARGOS, MINEROS, PFCMARGOS y PFGRUPOSURA fueron las acciones que presentaron pérdidas para el portafolio.

Figura 3-8: Ganancia neta por acción

Finalmente, la simulación del portafolio de esta sección utilizando diferentes modelos logra una mayor rentabilidad en comparación a las otras simulaciones donde solo se hacía uso de un tipo de modelo para todas las acciones. Debido a esto, se selecciona esta combinación, temporalmente, como la más acertada para el inversionista mientras se realizan otras combinaciones y se realiza la respectiva comparación de retorno sobre la inversión.

3.6.3.3 Selección de un subconjunto de acciones para el portafolio

Hasta el momento la mejor opción para el inversionista es utilizar un tipo de modelo diferente para cada acción según la mayor precisión balanceada alcanzada; con esto, en la simulación del portafolio se logra una rentabilidad de 36% (3,6 millones aproximadamente). Esta rentabilidad se genera con el consolidado de pérdidas o ganancias de las 27 acciones seleccionadas; sin embargo, como se observó en la Figura 3-8, tan solo dos acciones logran el 58% de la rentabilidad total, por lo que estos resultados sugieren la siguiente hipótesis: ¿un subconjunto de acciones podría generar un portafolio con mayores ganancias?. Para resolver esta pregunta se evaluaron diferentes portafolios que están conformados desde una hasta 27 acciones, para un total de 27 simulaciones. De esta manera, en esta sección se dará respuesta al Paso 4 de la metaheurística donde

se decide evaluar la inclusión gradual de las acciones, comenzando en una hasta llegar a 27 activos en el portafolio; esto se realiza con el fin de determinar qué número de acciones y cuáles son las más apropiadas para al inversionista al generar mayor rentabilidad. Para elegir qué acciones se van incluyendo, se opta por hacer uso de la ganancia por acción que se obtuvo en la simulación anterior. La Tabla 3-7 presenta dicha ganancia, donde la primera acción es la que tiene mayor rentabilidad, por lo que será el primer activo a incluir en el portafolio y se hará así sucesivamente hasta incluir las 27 acciones.

Tabla 3-7: Priorización de activos según ganancia para incluir en el portafolio.

Priorización	Activo	Ganancia	Costo transacción	Ganancia Neta
1	<i>ECOPETROL</i>	1.262.645	60.302	1.202.343
2	<i>CONCONCRET</i>	897.903	36.026	861.877
3	<i>GRUPOAVAL</i>	443.069	39.584	403.485
4	<i>ETB</i>	302.035	41.036	260.999
5	<i>PFBCOLOM</i>	302.920	59.947	242.973
6	<i>TERPEL</i>	240.990	21.534	219.456
7	<i>ISA</i>	266.100	46.785	219.315
8	<i>CORFICOLCF</i>	236.380	26.341	210.039
9	<i>CEMARGOS</i>	252.760	44.225	208.535
10	<i>PFGRUPOARG</i>	259.060	51.283	207.777
11	<i>PFAVAL</i>	237.710	35.725	201.985
12	<i>CELSIA</i>	191.590	30.209	161.381
13	<i>BOGOTA</i>	158.260	33.173	125.087
14	<i>PFDAVVNDA</i>	141.340	17.113	124.227
15	<i>BCOLOMBIA</i>	114.500	22.243	92.257
16	<i>PROMIGAS</i>	68.120	9.826	58.294
17	<i>GRUPOSURA</i>	91.100	47.865	43.235
18	<i>NUTRESA</i>	39.860	37.984	1.876
19	<i>CNEC</i>	-	-	-
20	<i>PFAVH</i>	-	-	-
21	<i>FABRICATO</i>	28.395	29.715	- 1.320
22	<i>GRUPOARGOS</i>	- 31.060	22.260	- 53.320
23	<i>EXITO</i>	- 44.060	32.557	- 76.617
24	<i>PFGRUPSURA</i>	- 119.100	20.860	- 139.960
25	<i>CLH</i>	- 210.605	30.204	- 240.809
26	<i>MINEROS</i>	- 332.120	26.730	- 358.850
27	<i>PFCEMARGOS</i>	- 324.790	43.834	- 368.624

Las señales de compra y venta utilizadas fueron las entregadas por los modelos en la simulación anterior, es decir, donde se eligió el mejor modelo por acción de acuerdo con el mejor desempeño predictivo en cuanto a la dirección del precio. En la Tabla 3-8 se muestra el resultado de las simulaciones de los 27 portafolios donde se puede observar la rentabilidad y la ganancia neta que genera cada uno de ellos. El tiempo de ejecución de estas 27 simulaciones fue de 30 minutos. De la tabla resalta que, para todos los portafolios, independientemente del número de acciones que lo conformen, la rentabilidad es positiva, lo cual es bastante beneficioso para el inversionista. Sin embargo, también es claro que la diferencia de rentabilidad de un portafolio a otro es bastante notoria, pues hay rentabilidades desde 34% hasta 78%, lo cual sugiere que el número de acciones a incluir en el portafolio si tiene influencia en la rentabilidad para el inversionista.

Tabla 3-8: Rentabilidad del portafolio usando diferente número de acciones.

# Acciones	ECOPETROL	CONCRETO	GRUPOAVAL	ETB	PFCOLOM	TERPEL	ISA	CORFICOLCF	CEMARGOS	PFCRUPOARG	PFAVAL	CELSIA	BOGOTA	PFCDAVVINDA	ECOLOMBIA	PFCMIGAS	GRUPOSURA	NUTRESA	CNEC	PFAVH	FABRICATO	GRUPOARGOS	EXITO	PFCRUPOPSURA	CLH	MINEROS	PFCMARGOS	Ganancia neta	Rentabilidad		
1	X																											3.420.390	34,2%		
2	X	X																											5.205.956	52,1%	
3	X	X	X																										5.877.669	58,8%	
4	X	X	X	X																									7.137.454	71,4%	
5	X	X	X	X	X																								6.880.546	68,8%	
6	X	X	X	X	X	X																							6.625.381	66,3%	
7	X	X	X	X	X	X	X																						7.268.364	72,7%	
8	X	X	X	X	X	X	X	X																					7.186.365	71,9%	
9	X	X	X	X	X	X	X	X	X																				5.988.918	59,9%	
10	X	X	X	X	X	X	X	X	X	X																			6.832.060	68,3%	
11	X	X	X	X	X	X	X	X	X	X	X																		7.850.457	78,5%	
12	X	X	X	X	X	X	X	X	X	X	X	X																	7.022.438	70,2%	
13	X	X	X	X	X	X	X	X	X	X	X	X	X																7.156.349	71,6%	
14	X	X	X	X	X	X	X	X	X	X	X	X	X	X															6.448.973	64,5%	
15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X														6.686.141	66,9%	
16	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X													6.829.889	68,3%	
17	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X												6.863.296	68,6%	
18	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X											5.848.614	58,5%	
19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X										5.848.614	58,5%	
20	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X									5.848.614	58,5%	
21	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								5.621.351	56,2%	
22	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X							5.261.438	52,6%	
23	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X						4.878.229	48,8%	
24	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X					4.513.013	45,1%	
25	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				4.184.669	41,8%	
26	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			3.708.638	37,1%	
27	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		3.605.640	36,1%	

De la tabla anterior se observa que la inclusión de acciones no presenta una dinámica en particular; sin embargo, se puede apreciar que el punto donde se alcanza la mayor rentabilidad del portafolio es al incluir 11 acciones, y también se observa que a partir de incluir 18 o más acciones en el portafolio hay una tendencia a la baja de la rentabilidad. De

esta manera, se puede seleccionar entonces 11 acciones como la conformación del portafolio más conveniente para el inversionista y, este caso, estaría conformado por:

- ECOPETROL
- CONCRET
- GRUPOAVAL
- ETB
- PFBCOLOM
- TERPEL
- ISA
- CORFICOLCF
- CEMARGOS
- PFGRUPOARG
- PFAVAL

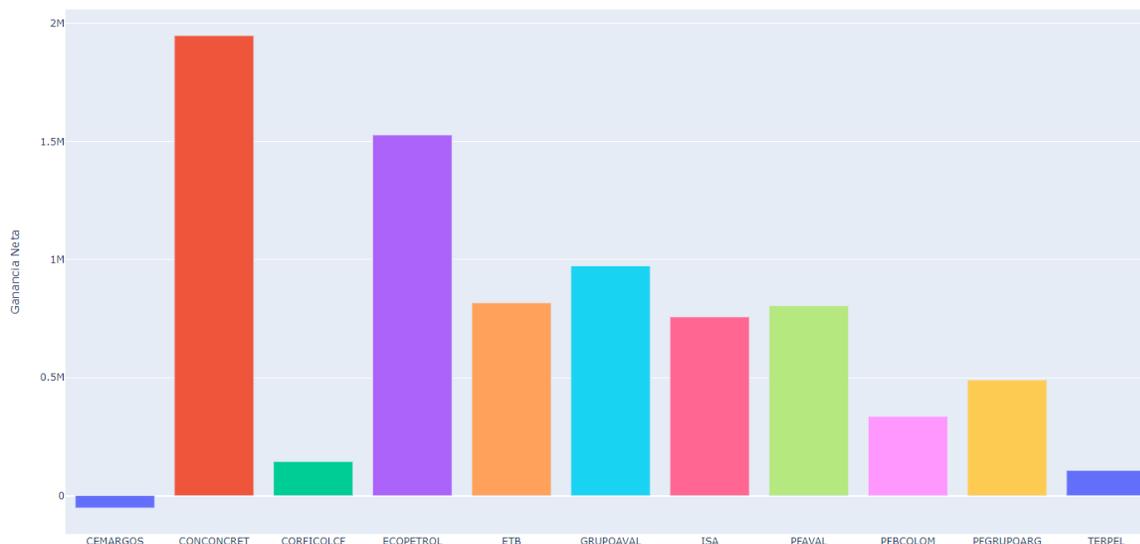
Hasta ese punto de la metaheurística lo más rentable para el inversionista es pronosticar las señales de compra y venta de las acciones de acuerdo al mejor tipo de modelo por acción, el cual es seleccionado por la mayor precisión balanceada. Adicionalmente, la conformación del portafolio sugiere que debe hacerse con 11 de las 27 acciones seleccionadas previamente. Estas reglas logran una rentabilidad de 78,5% (7,8 millones) para el inversionista. En la Figura 3-9 se presenta la evolución en el tiempo del portafolio conformado por 11 acciones, donde se aprecia una clara tendencia alcista durante casi todo el periodo de la simulación.

Figura 3-9: Evolución portafolio con 11 acciones.



Por otra parte, en la Figura 3-10 se presenta la ganancia neta por acción, donde sobresale CONCRET como el activo con mayor rentabilidad, lo cual difiere de la mejor simulación del Paso 4 de la metaheurística, donde la acción más rentable era ECOPETROL (ver Figura 3-8). Adicionalmente se aprecia que CEMARGOS es la única acción que genera pérdidas para el portafolio, aunque su valor no es muy significativo.

Figura 3-10: Ganancia por acción del portafolio conformado por 11 acciones.



Finalmente, a continuación, se detallará el último paso de la metaheurística donde se elegirá la mejor combinación de reglas que conforman la estrategia de trading automático, y que entregan la más alta rentabilidad evaluada para el inversionista.

3.6.3.4 Evaluación del porcentaje de capital a invertir en cada operación

Para las simulaciones anteriores, se fijó en 40% el porcentaje del capital disponible a invertir en cada operación. Sin embargo, en esta sección se propone realizar un análisis de sensibilidad y desarrollar el Paso 5 de la metaheurística donde se evaluarán diferentes porcentajes de capital a invertir en cada movimiento, con el fin de determinar el valor que más favorezca al inversionista. Este parámetro es importante dado que es una manera de regular la diversificación del portafolio, pues si su valor es cercano a 100% indica que todo

el capital disponible se invertirá en una única acción, lo cual deposita todo el riesgo en dicho activo. Para esto, se propone evaluar los siguientes porcentajes: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% y 100%.

En la Tabla 3-9 se presentan los resultados de las simulaciones de los 10 portafolios teniendo en cuenta el porcentaje del capital disponible a invertir en cada operación. En la tabla se observa que la rentabilidad es directamente proporcional a dicho porcentaje, es decir, entre más alto sea el porcentaje a transar en cada operación, más rentabilidad generará el portafolio. Esto sugiere que tener una o pocas acciones al tiempo en el portafolio es más rentable que tener muchos activos diferentes; sin embargo, en este punto es importante tener en cuenta la aversión al riesgo del inversionista ya que la diversificación de la cartera ha sido reconocida como una herramienta de inversión útil que puede reducir efectivamente el riesgo de incertidumbre futura en los mercados financieros (Eom, Kaizoji, Livan & Scalas, 2021).

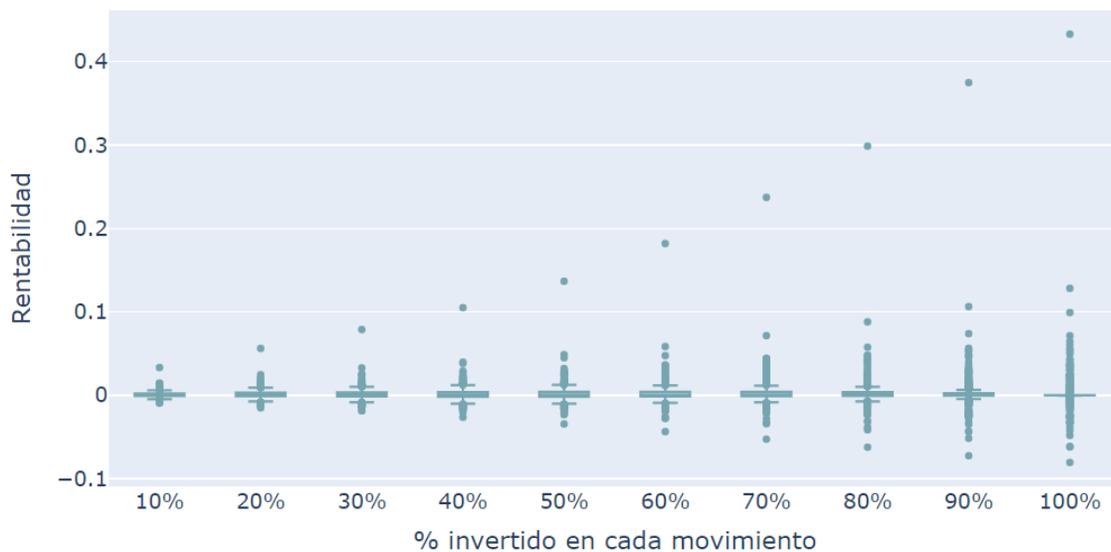
Tabla 3-9: Simulación portafolio con diferente porcentaje a invertir en cada operación.

% invertir en cada operación	Ganancia neta	Rentabilidad
100%	19.457.330	195% 
90%	17.459.586	175% 
80%	15.637.730	156% 
70%	13.469.290	135% 
60%	11.451.931	115% 
50%	9.909.753	99% 
40%	7.850.457	79% 
30%	6.196.111	62% 
20%	4.489.236	45% 
10%	2.560.076	26% 

De acuerdo con lo anterior y con el fin de realizar un estimativo del riesgo en que incurriría el inversionista en cada uno de los escenarios, se preparó la Figura 3-11 y la Figura 3-12. Por una parte, en la Figura 3-11 se presenta, para cada portafolio, un diagrama de caja con las rentabilidades durante todo el periodo de simulación que permite ver la pérdida y

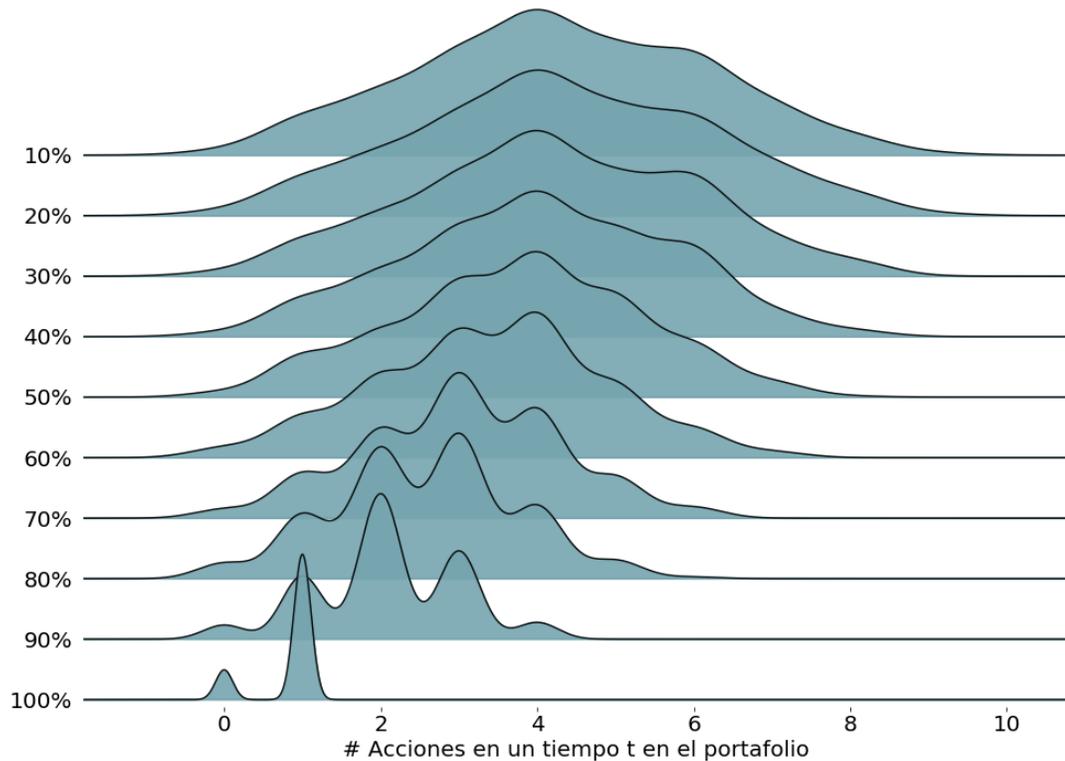
ganancia máxima alcanzada. A partir de esto, se observa que, para los rendimientos, a medida que aumenta el porcentaje a invertir en cada operación también aumenta la dispersión de dichos rendimientos, lo cual implica más volatilidad. Se ve que, para pequeños porcentajes a invertir en cada movimiento, hay pequeñas pérdidas, pero también las rentabilidades positivas son de magnitud baja. Por ejemplo, para porcentajes como 10% y 20%, la pérdida máxima alcanzada es tan solo del 1%, pero la ganancia máxima también se restringe y se aproxima a 5%. Caso contrario ocurre cuando se decide invertir todo en cada operación (100%), ya que se observa que hay rendimientos muy negativos, incluso alcanzando el 10% de pérdida; sin embargo, también hay otros puntos de recuperación con ganancias superiores a 40%.

Figura 3-11: Rentabilidad de los portafolios según porcentaje invertido en cada operación



Por otra parte, la Figura 3-12 presenta, para los diferentes porcentajes a invertir en cada movimiento, el número de acciones que van conformando el portafolio a lo largo de los días del periodo de simulación, puesto que como se ha mencionado, cada día pueden existir un número diferente de acciones en la cartera. De esta figura se puede decir que: entre más bajo sea el porcentaje a invertir, más activos diferentes conforman la cartera, lo cual puede reducir el riesgo al tener diversificado el portafolio. Esta diversificación no se da únicamente en el último caso, al invertir el 100% en cada movimiento, ya que todo el tiempo solo hay una acción en el portafolio o ninguna.

Figura 3-12: Número de acciones de los portafolios según porcentaje invertido en cada operación



Según el análisis anterior, se concluye que, entre más alto sea el porcentaje invertido en cada movimiento, más riesgo inherente habrá; esto sucede ya que el portafolio no estará diversificado, sino que siempre dependerá de la rentabilidad de una única acción. Adicionalmente, para este escenario, los rendimientos presentan una alta volatilidad, lo que indica que, así como se pueden presentar grandes ganancias también podrán ocurrir pérdidas cuantiosas. De esta manera, en este paso de la metaheurística se deja a consideración del inversionista y su aversión al riesgo, seleccionar ese porcentaje que invertirá en cada movimiento. Sí el porcentaje seleccionado es bajo indica que el inversionista tiene un perfil conservador, donde busca conservar su capital y tolera bajos pero estables rendimientos. De otro lado, sí el porcentaje está en el medio, refleja que el inversionista está en búsqueda de rendimientos razonables para su inversión. Finalmente, sí el porcentaje seleccionado es alto, indica que el inversionista tiene un perfil agresivo, y desea altos retornos y tolera pérdidas considerables en su inversión.

En este sentido, para efectos prácticos de este trabajo se selecciona el porcentaje a invertir en cada operación en 40%, tal como se había definido inicialmente para las simulaciones anteriores. Con este valor, la estrategia de trading automático construida genera un rendimiento de 79% aproximadamente. Con esto, se procede a comparar dicha estrategia con la estrategia tradicional de conformación de portafolio de mínima varianza, la cual se basa en el modelo de Harry Markowitz, en el que relaciona riesgo y rentabilidad. Esta comparación se hace con el objetivo de evaluar la estrategia de trading algorítmico frente a una estrategia tradicional y, de esta manera determinar si hay o no un valor agregado en la predicción de señales de compra y venta que entregan los modelos de inteligencia artificial.

3.6.4 Comparación de la estrategia trading algorítmico vs. la estrategia mínima varianza

Como se mencionó anteriormente, en esta sección se hace la simulación del portafolio de mínima varianza, el cual está basado en el modelo propuesto por Markowitz, donde los portafolios se conforman basados en el riesgo (variabilidad del retorno) y el rendimiento esperado. En este escenario, los inversionistas buscan rendimientos con la menor variabilidad posible, es decir, tienen aversión al riesgo. En este modelo de Markowitz se busca una diversificación eficiente, la cual no está basada en el número de activos diferentes, sino que está basada en la correlación de las acciones (Medina, 2003). Es de resaltar que la estrategia de Markowitz en su versión más pura no pronostica los movimientos de precio de los activos.

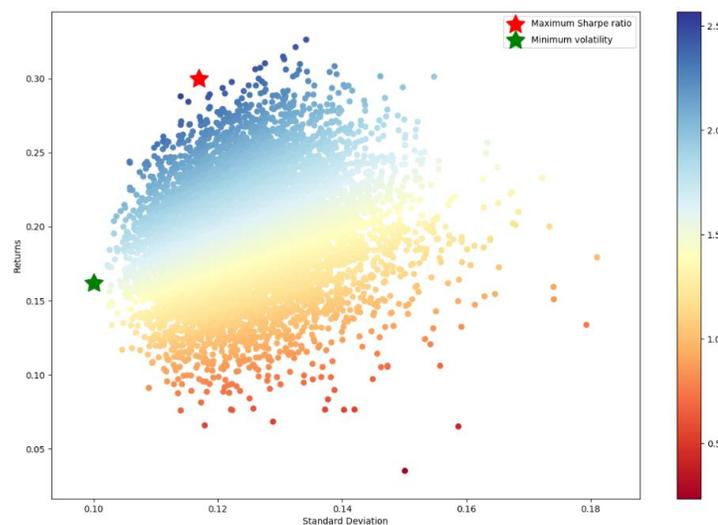
La simulación del portafolio de mínima varianza se realiza con el objetivo de comparar su rendimiento alcanzado con el de la estrategia de trading automático, que, bajo las condiciones actuales, el retorno fue de 79% aproximadamente. A partir de esta comparación se podrá estimar que es más conveniente para el inversionista, sí llevar a cabo una estrategia tradicional o incursionar en otro tipo de estrategias automáticas generadas por señales de compra y venta de modelos de machine learning.

3.6.4.1 Supuestos para la conformación de portafolio según modelo Markowitz:

Para hacer las dos estrategias comparables (Markowitz y trading automático), se plantean los siguientes supuestos:

- Para definir el porcentaje de capital total a invertir cada día en el portafolio de Markowitz (suma de compra de todas las acciones), se toma como referencia el porcentaje invertido cada día en el portafolio de la estrategia de trading automático. Pues si bien en esta estrategia el porcentaje de cada movimiento era 40%, en un día se podían ejecutar diversas operaciones ya que se podían comparar varias acciones.
- La reconformación del portafolio se hará de forma diaria ya que así se realizó en el portafolio con la estrategia de trading automático.
- Se realizan 10.000 conformaciones de portafolios diferentes cada día con el fin de seleccionar la mejor participación de acciones según la menor varianza alcanzada. En otras palabras, se construye la frontera eficiente con 10.000 puntos diferentes con su varianza y rendimiento respectivo, y se selecciona cada día el punto con el rendimiento que logra la menor variabilidad (estrella verde en la Figura 3-13).

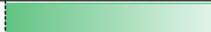
Figura 3-13: Ejemplo frontera eficiente



3.6.4.2 Resultados obtenidos

En la Tabla 3-10 se presenta la ganancia y rentabilidad obtenida con las estrategias de trading automático vs. Markowitz. En dicha tabla se observa que mientras la estrategia algorítmica logra una rentabilidad positiva de 79%, la estrategia de mínima varianza obtiene pérdidas aproximadamente del 5%.

Tabla 3-10: Rentabilidad portafolio con estrategia de trading automático vs. Markowitz

Estrategia	Ganancia neta	Rentabilidad
Trading Automático	7.850.457	79% 
Mínima varianza (Markowitz)	- 511.696	-5% 

Asimismo, en la Figura 3-14 se presenta la evolución del valor del portafolio obtenido con la estrategia de trading automático (línea roja) vs. la estrategia de mínima varianza – Markowitz (línea azul). Se observa que la estrategia de Markowitz inclusive alcanza pérdidas superiores a la final obtenida, pues llega a valores del portafolio de 8 millones, lo cual significa una pérdida de 20% del valor inicial del capital. Caso contrario sucede con el valor del portafolio obtenido con la estrategia algorítmica de negociación, donde solo se presentan pequeñas variaciones negativas del capital invertido en los primeros meses, y posteriormente se ve una marcada tendencia alcista, alcanzando al final del periodo de simulación una rentabilidad de 79%.

Figura 3-14: Valor del portafolio con estrategia de trading automático vs. Markowitz

Según lo anterior, queda evidenciado que para el inversionista es más rentable seguir una estrategia de negociación basada en trading automático, pues la rentabilidad obtenida es muy superior a la de la estrategia tradicional de conformación de portafolio, donde en este caso hay pérdidas. Así pues, los resultados muestran que las predicciones de las señales de compra y venta que arrojan los modelos a partir del entrenamiento del movimiento del precio son efectivas para la conformación del portafolio, dado que en las diferentes simulaciones que se hicieron nunca hubo pérdidas, sino que se observó rentabilidades bastantes buenas.

3.7 Conclusiones

En este capítulo se mostraron los resultados de diversas estrategias de negociación algorítmicas a través de un modelo simulación de portafolio. Dicho modelo mostró las bondades que tiene para el inversionista hacer uso de trading algorítmico, pues al evolucionar la estrategia, se logró una rentabilidad de 78%. Estos resultados fueron contrastados con la rentabilidad que genera la estrategia tradicional de conformación de portafolio – Markowitz-, donde esta estrategia generó una pérdida de 5% sobre la inversión. En este sentido, se evidencian los beneficios de construir una estrategia basada en el pronóstico de la tendencia del precio con el fin de conformar el portafolio.

Adicionalmente, con respecto a los tipos de modelos de pronóstico utilizados para la tendencia del precio, se encontró que la mejor alternativa es seleccionar el tipo de modelo que mejor desempeño predictivo tiene para cada activo. Esto se sustenta en el hecho de que se probó usar una misma técnica para todas las acciones, lo cual resultó no ser tan rentable comparado con usar las señales de compra y venta de las acciones provenientes de las predicciones del mejor modelo para cada activo. En este punto, aunque no haya un tipo de modelo que sea consistentemente superior para todas las acciones, la regresión logística resaltó como el tipo de modelo que mejor predice la tendencia del precio para casi la mitad de los activos seleccionadas (48%).

Finalmente, otro hallazgo que se obtuvo de la construcción de la estrategia de negociación algorítmica es que es más rentable seleccionar un subconjunto de acciones para conformar el portafolio, en lugar de seleccionar todos los activos que están en el COLCAP. Particularmente, en la estrategia construida en este trabajo, el subconjunto que más rentabilidad le brinda al inversionista es un total de 11 acciones.

4. Conclusiones

4.1 Respuesta a las preguntas de investigación

En el Capítulo 1 se plantó la siguiente pregunta de investigación:

¿Cuáles metodologías de Inteligencia Artificial son óptimas para aplicar en el mercado financiero colombiano con el fin de generar algoritmos de negociación que generen rentabilidad?

Como respuesta se puede decir lo siguiente:

En el desarrollo de este trabajo se hicieron uso de diferentes metodologías de inteligencia artificial, tales como: regresión logística, árboles de decisión, bosques aleatorios, vecinos más cercanos y red neuronal recurrente. Por un lado, se compararon dichas técnicas con base en el desempeño predictivo del movimiento del precio de las acciones y, por otro, se usaron dichas predicciones como señales de compra y venta para tomar decisiones de negociación, a través de lo cual se pudo obtener la rentabilidad para el inversionista al conformar el portafolio con base en dichos modelos de pronóstico. A partir de esta comparación, se encontró que no existe una metodología que sea consistentemente superior ni inferior para predecir la dirección del precio de todas las acciones. Por el contrario, los resultados muestran que, para los distintos activos seleccionados del mercado financiero colombiano, hay metodologías de IA que favorecen más a unas acciones que a otras.

Con los resultados anteriores, la recomendación es no usar un único tipo de modelo para generar las señales de compra y venta, sino que lo más rentable de cara al inversionista es encontrar el modelo que más favorezca a cada activo según su desempeño predictivo. Sin embargo, es importante afirmar que, sí hay algunas metodologías que sobresalen más que otras; esto se sustenta al obtener que, de las 5 técnicas comparadas, para el 48% aproximadamente de las acciones estudiadas, la mejor metodología resultante fue la regresión logística. Si bien esta metodología no es superior para todas las acciones, si lo hace para la mitad de ellas, lo que hace resaltar su poder al momento de predecir la dirección del precio. En segundo lugar,

y a diferencia de la regresión logística, se encontró un método no paramétrico que logra ser superior en el 22% de los activos seleccionados: vecinos más cercanos.

En este sentido, y aunque existan algunas técnicas que resaltan más, como regresión logística y vecinos más cercanos, la conclusión es que, para conformar el portafolio con distintas acciones, la mejor estrategia de trading algorítmico es usar las predicciones del movimiento del precio a partir del mejor tipo de modelo que se obtenga para cada acción. Esto se debe en parte a que el comportamiento y la dinámica de las series de los activos es diferente, por lo que no hay un único tipo de modelo que pueda representarlas a todas, sino que dependiendo de su estructura hay técnicas que pueden favorecer más la predicción.

4.2 Cumplimiento de los objetivos

4.2.1 Objetivo general

El objetivo general propuesto en esta tesis fue:

Construir un prototipo de simulación para evaluar estrategias de Comercio Algorítmico en el mercado financiero colombiano, con el fin de analizar si es posible incorporar este tipo de estrategias por parte de los inversionistas.

Para el desarrollo de este objetivo se tomó en consideración lo siguiente:

En la Sección 3.6.1 se presentó el esquema con el modelo de simulación propuesto, el cual consistió en 4 fases: ingestión de datos, entrenamiento de modelos, simulación de portafolio y cálculo de indicadores. Estos 4 pasos fueron programados en Python, y como tal la consecución de éstos se realiza de forma automática y articulando un paso con otro. Solo basta con que el inversionista establezca algunos parámetros para llevar a cabo la estrategia de acuerdo a su aversión de riesgo (pérdida máxima aceptada y porcentaje a invertir en cada operación). En este sentido, el modelo de simulación comienza con la extracción de los datos de la BVC, luego entrena un modelo de IA para cada activo, posteriormente realiza la simulación de portafolio donde se puede apreciar de forma diaria el valor del portafolio y finalmente calcula indicadores como la rentabilidad final obtenida.

Para la segunda fase del modelo de simulación, dado que en el sitio web de la Bolsa de Valores de Colombia están disponibles los datos históricos del mercado de capitales de renta variable, fue posible extraer parte de esta información para entrenar modelos predictivos del movimiento del precio de las acciones y, otra parte para realizar la respectiva evaluación de los modelos a través de un backtesting. Para dicho backtesting, se generaron las predicciones de la dirección del precio de los diferentes activos mediante los modelos ya entrenados, con el fin de convertirlas en señales de negociación para incluir en la estrategia de trading algorítmico. A partir de estas señales de compra y venta se construyeron diversas estrategias, las cuales diferían del tipo de modelo usado para la predicción de la dirección del precio de las acciones, del subconjunto de activos seleccionados y del capital invertido en cada movimiento.

Según lo anterior, la estrategia de trading automático fue evolucionando gradualmente con el objetivo de obtener la máxima ganancia posible para el inversionista; la estrategia finalmente seleccionada se comparó con el rendimiento que genera la estrategia tradicional de conformación de portafolio: mínima varianza o Markowitz. El hallazgo a partir de dicha comparación mostró que la rentabilidad que genera la estrategia de trading algorítmico es muy superior a la estrategia tradicional, donde inclusive esta última genera pérdidas, mientras que la primera supera rentabilidades del 78%. Finalmente, esta simulación evidencia que es bastante rentable para el inversionista incorporar estrategias automáticas de negociación basadas en predicciones de modelos de IA, pues los resultados indicaron que dichas estimaciones logran generar señales de compra y venta pertinentes para la conformación automática del portafolio al generar altas utilidades para el inversionista.

4.2.2 Objetivo Específico 1

El primer objetivo específico planteado fue:

Proponer una estrategia para desarrollar modelos que permitan pronosticar el movimiento de precios para las principales acciones en el mercado financiero colombiano.

La estrategia fue la siguiente:

Las series de tiempo de las acciones fueron transformadas en problemas de clasificación, en donde la variable respuesta se definió como la dirección del precio del día siguiente (sube el valor o no), por lo cual la temporalidad de la predicción se estableció como diaria. En ese sentido, cada día se evaluaba si el precio de ese día era superior o inferior al precio del día siguiente. En caso de ser inferior, la variable respuesta tomaba un valor de 1 ya que la tendencia del precio indicaba un alza el día siguiente, y en caso de ser superior la variable respuesta era 0 ya que la tendencia sería bajista.

Para entrenar los modelos de clasificación, se consideraron inicialmente las siguientes variables explicativas: índice COLCAP, ETF SPY y precio de cierre, precio mayor, precio menor, volumen y cantidad negociada de cada acción; sin embargo, a través de una metodología propuesta se realizó la respectiva selección de características. Con este planteamiento, se seleccionaron 5 metodologías diferentes de modelos de clasificación con el fin de comparar su desempeño predictivo y de esta manera seleccionar la (s) técnica (s) que mejor predice (n) la dirección del precio de las acciones. Para cada activo se entrenaron los 5 tipos de modelos seleccionados y a través de la medida “precisión balanceada” se comparó su rendimiento.

De acuerdo con el contexto lo anterior, se encontró que no hay un tipo de modelo que sea consistentemente superior ni inferior para predecir el movimiento del precio de todas las acciones, es decir, para cada acción el mejor modelo variaba entre un activo y otro. Un punto importante para resaltar es que, para todos los modelos entrenados, la precisión balanceada oscilaba alrededor del 50%; esto es debido en parte a la alta volatilidad que presentan los precios de las acciones, pues existen diversos factores externos que influyen en las fluctuaciones que a diario presentan dichos activos, por ejemplo, cambios inesperados en la política monetaria, eventos geopolíticos, quiebras bancarias, entre otros. Sin embargo, simplemente con analizar la precisión balanceada no es posible saber si las señales de compra y venta que generan los modelos serán rentables o no para el inversionista, por lo que este análisis se complementó con la utilidad o pérdida que generan estas predicciones al momento de usarlas para tomar decisiones y conformar el portafolio, lo cual se desarrolló para dar cumplimiento al Objetivo Específico 2.

4.2.3 Objetivo Específico 2

El segundo objetivo específico planteado fue:

Adaptar una estrategia para la construcción de portafolios que incorpore los modelos de pronósticos desarrollados en el objetivo anterior.

Para cumplir con este objetivo se hizo uso de las predicciones resultantes en el desarrollo del Objetivo Específico 1, las cuales se obtuvieron a partir de los diferentes modelos entrenados para cada una de las acciones. Dichas predicciones indicaban si el precio subiría o no el día siguiente, por lo cual estos pronósticos del movimiento del precio fueron convertidos en señales de compra y venta de los activos con el fin de construir una estrategia para la conformación de portafolio.

En este contexto, para definir la estrategia, cuando la predicción indicaba que el precio subiría el día siguiente, la señal asociada era comprar, pues todo apuntaba que el inversionista podría tener rentabilidad al comprar a un precio inferior de lo que estaría en el futuro cercano. En contraste, cuando la predicción reflejaba que el precio no subiría el día siguiente surgían dos escenarios, uno es cuando hubieran acciones en el portafolio, donde la señal indicaba que el inversionista debería vender para evitar pérdidas, y el segundo escenario es cuando no había acciones de dicho activo en el portafolio, lo cual indicaría que el inversionista no debería realizar movimientos de dicha acción.

Según las reglas anteriores, se utilizaron las predicciones para construir una estrategia automática de negociación que indicaba las señales y momentos de compra y venta de las acciones, lo cual permitió la reconfiguración diaria del portafolio. Con esto, fue posible realizar la simulación del portafolio, donde se pudo apreciar el comportamiento de la cartera lo largo del periodo de tiempo establecido. En dicha simulación se observó una tendencia alcista, donde el valor del portafolio iba incrementando hasta alcanzar una rentabilidad del 78% para el inversionista, lo cual pone en evidencia que implementar este tipo de estrategias automáticas

basadas en los pronósticos de modelos de IA es bastante positivo para los inversores que le apuestan al mercado de capitales de renta variable.

4.2.4 Objetivo Específico 3

El tercer objetivo específico planteado fue:

Validar el modelo desarrollado en los Objetivos 1 y 2 mediante un modelo de simulación el cual permita especificar estrategias para decisiones de conformación de portafolio.

A partir de las predicciones transformadas en señales de compra y venta, para el desarrollo de este objetivo se realizó un modelo de simulación, el cual se automatizó y como tal el inversionista puede cambiar, de acuerdo a su apetito de riesgo, algunos parámetros como el porcentaje a invertir en cada movimiento, la pérdida máxima soportada, entre otros, con el fin de adecuar la estrategia de negociación y realizar la respectiva simulación para obtener la rentabilidad. Para probar dicho modelo, se tomó un periodo de tiempo a modo de backtesting en el que se realizó la conformación diaria del portafolio con base en las señales entregadas por los modelos de pronóstico. Es importante resaltar que se realizaron varias simulaciones del portafolio con base en modificaciones que se hacían sobre la estrategia de trading algorítmico, la cual buscaba maximizar la rentabilidad. Entre las modificaciones estaba el tipo de modelo que generaba las predicciones, tomar un subconjunto de acciones para formar el portafolio, etc.

La estrategia final consistió en un subconjunto de 11 acciones y, señales generadas por el mejor modelo para cada acción; al realizar la simulación se generó una rentabilidad de 78% sobre la inversión, lo cual representa una cifra bastante favorable para el inversionista. Dicha estrategia fue comparada con la rentabilidad generada por la estrategia tradicional de conformación de portafolio: mínima varianza, para lo cual también se hizo una simulación para dicha estrategia dando como resultado para el inversionista una pérdida del 5%.

De acuerdo con lo descrito en el párrafo anterior, se puede concluir que implementar una estrategia de trading automático basado en la predicción del movimiento del precio de las acciones es más rentable que llevar a cabo una estrategia de inversión basada en la teoría tradicional de mínima varianza, la cual no considera el pronóstico del precio de las acciones. De esta manera, el trading algorítmico permite al inversionista tener un panorama futuro de lo que sucederá con el precio de las acciones, lo cual le permite anticiparse y tomar acciones en pro de su rentabilidad.

Bibliografía

- Aloud, M. E. (2020). The role of attribute selection in Deep ANNs learning framework for high-frequency financial trading. *Intelligent Systems in Accounting, Finance and Management*, 27(2), 43-54.
- Ardanuy, J. (2012). Breve introducción a la bibliometría. Universidad de Barcelona.
- Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert systems with Applications*, 36(7), 10696-10707.
- Azzini, A & Tettamanzi, A. (2008). Evolutionary single-position automated trading. *Workshops on Applications of Evolutionary Computation*, Springer, Berlin, Heidelberg, pp. 62-72.
- Bahadur, L. R. (2015). Financial Literacy: The Indian Story. *World Journal of Social Sciences*, 5(3), 45-57.
- Banco de la República. (s.f.). *Mercado accionario*. <https://www.banrep.gov.co/es/estadisticas/mercado-accionario>
- Barbosa, R. P., & Belo, O. (2008). Algorithmic Trading Using Intelligent Agents. *In IC-AI* (pp. 136-142).
- BBC (2018a). *Wall Street: qué hay detrás de la estrepitosa caída de la bolsa en el último mes que borró las ganancias de todo 2018*. <https://www.bbc.com/mundo/noticias-45983822>
- BBC (2018b). *Wall Street sufre su peor semana en 10 años (qué significa para la economía global)*. <https://www.bbc.com/mundo/noticias-46651665>
- Berutich, J. M., López, F., Luna, F., & Quintana, D. (2016). Robust technical trading strategies using GP for algorithmic portfolio selection. *Expert Systems with Applications*, 46, 307-315.

- Berutich, J. M., López, F., Luna, F., & Quintana, D. (2016). Robust technical trading strategies using GP for algorithmic portfolio selection. *Expert Systems with Applications*, 46, 307-315.
- Blitz, D., & Huij, J. (2012). Evaluating the performance of global emerging markets equity exchange-traded funds. *Emerging markets review*, 13(2), 149-158.
- Bolsa de Valores de Colombia (BVC) (2011). *METODOLOGÍA PARA EL CÁLCULO DEL ÍNDICE COLCAP*.
https://www.bvc.com.co/recursos/Files/Mercados/Indices_Accionarios/Nueva_Metodologia_Calculo_Indice_COLCAP.pdf
- Bolsa de Valores de Colombia (BVC) (2020). *Noticias Generales*.
https://www.bvc.com.co/pps/tibco/portalbvc/Home/NoticiasDetalle?com.tibco.ps.pagesvc.renderParams.sub5d9e2b27_11de9ed172b_-783b7f000001=rp.docURI%3Dpof%253A%252Fcom.tibco.psx.model.cp.Document%252F6ed219b4_1700ab5155b_-4a2ac0a84ca9%26rp.currentDocumentID%3D6ed219b4_1700ab5155b_-4a2ac0a84ca9%26action%3DopenDocument%26addDefaultTarget%3Dfalse%26
- Bolsa de Valores de Colombia (BVC) (s.f.-a). *Diccionario de Finanzas*.
<https://www.bvc.com.co/pps/tibco/portalbvc/Home/Glosario>
- Bolsa de Valores de Colombia (BVC) (s.f.-b). *Mercado de Renta Variable*.
<https://www.bvc.com.co/pps/tibco/portalbvc/Home/Mercados/enlinea/acciones?action=dummy>
- Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8), 2267-2306.
- Brogaard, J., Hendershott, T., & Riordan, R. (2017). High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics*, 124(1), 22-42.
- Brogaard, J., Hendershott, T., Hunt, S., & Ysusi, C. (2014). High-frequency trading and the execution costs of institutional investors. *Financial Review*, 49(2), 345-369.
- Brownlees, C. T., Cipollini, F., & Gallo, G. M. (2010). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, 9(3), 489-518.
- Cartea, A., & Jaimungal, S. (2013). Modelling asset prices for algorithmic and high-frequency trading. *Applied Mathematical Finance*, 20(6), 512-547.
- Cartea, Á., Jaimungal, S., & Kinzebulatov, D. (2016). Algorithmic trading with learning. *International Journal of Theoretical and Applied Finance*, 19(04), 1650028.

- Cartlidge, J., Szostek, C., De Luca, M., & Cliff, D. (2012). Too Fast Too Furious-Faster Financial-market Trading Agents Can Give Less Efficient Markets. *In ICAART (2)* (pp. 126-135).
- Chakole, J. B., Kolhe, M. S., Mahapurush, G. D., Yadav, A., & Kurhekar, M. P. (2021). A Q-learning agent for automated trading in equity stock markets. *Expert Systems with Applications*, 163, 113761.
- Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901-923.
- Choudhry, R., & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39(3), 315-318.
- Chu, J., Chan, S., & Zhang, Y. (2020). High frequency momentum trading with cryptocurrencies. *Research in International Business and Finance*, 52, 101176.
- Cordón, O., Herrera, F., Holfmann, F., & Magdalena, L. (2001). *Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases* (Vol. 19). World Scientific.
- Dejavite, C. H & Lima, P. A. (2010). Towards automated trading based on fundamentalist and technical data. *Brazilian Symposium on Artificial Intelligence*, Springer, Berlin, Heidelberg, pp. 112-121.
- Ding, S., Hanna, J., & Hendershott, T. (2014). How slow is the NBBO? A comparison with direct exchange feeds. *Financial Review*, 49(2), 313-332.
- Driaunys, K., Masteika, S., Sakalauskas, V., & Vaitonis, M. (2014). An algorithm-based statistical arbitrage high frequency trading system to forecast prices of natural gas futures. *Transformations in Business & Economics*, 13(3), 96-109.
- Eom, C., Kaizoji, T., Livan, G., & Scalas, E. (2021). Limitations of portfolio diversification through fat tails of the return Distributions: Some empirical evidence. *The North American Journal of Economics and Finance*, 56, 101358.
- Farjam, M., & Kirchkamp, O. (2018). Bubbles in hybrid markets: How expectations about algorithmic trading affect human trading. *Journal of Economic Behavior & Organization*, 146, 248-269.
- Fister, D., Mun, J. C., Jagrič, V., & Jagrič, T. (2019). Deep Learning for Stock Market Trading: A Superior Trading Strategy?. *Neural Network World*, 29(3), 151-171.
- Fukuma, N., & Kadogawa, Y. (2020). An Overview of Algorithmic Trading in Foreign Exchange Markets and Its Impacts on Market Liquidity. *Bank of Japan Review Series*,

(20-E-5).

- Gabrielsson, P., & Johansson, U. (2015, December). High-frequency equity index futures trading using recurrent reinforcement learning with candlesticks. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 734-741). IEEE.
- Gloukhov, V. V., Ilin, I. V., Kuposov, V. I., & Levina, A. I. (2014). Market risk neutral strategies: modeling and algorithmization. *Asian Social Science*, *10*(24), 209.
- Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*, *23*(8), 800-808.
- Hendershott, T., & Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis*, *48*(4), 1001-1024.
- Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity?. *The Journal of finance*, *66*(1), 1-33.
- Hi'ovská, K., & Koncz, P. (2012). Application of Artificial Intelligence and Data Mining Techniques to Financial Markets. *Economic Studies & Analyses/Acta VSFS*, *6*(1), 62-76.
- Hogenboom, F. (2012). Financial events recognition in web news for algorithmic trading. *International Conference on Conceptual Modeling*, Springer, Berlin, Heidelberg, pp. 368-377.
- Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E & Liu. M. (2015). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, *36*, 534-551.
- Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, *32*(10), 2513-2522.
- Janosek, M., Volna, E., & Kotyrba, M. (2015). Knowledge discovery in dynamic data using neural networks. *Cluster Computing*, *18*(4), 1411-1421.
- Kablan, A. (2009, October). Adaptive neuro fuzzy inference systems for high frequency financial trading and forecasting. In *2009 Third International Conference on Advanced Engineering Computing and Applications in Sciences* (pp. 105-110). IEEE.
- Kayal, A. (2010, October). A neural networks filtering mechanism for foreign exchange trading signals. In *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on* (Vol. 3, pp. 159-167). IEEE.

- Khoury, P., & Gorse, D. (2015). Investing in emerging markets using neural networks and particle swarm optimization. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.
- Kissell, R. L. (2013). *The science of algorithmic trading and portfolio management*. Academic Press, pp. 1-473.
- Kong, X. (2013). Study of cep based on dynamic data management system and algorithmic trading. *Journal of Theoretical & Applied Information Technology*, 48(2).
- Kunz, K., & Martin, J. (2015). Into the breach: the increasing gap between algorithmic trading and securities regulation. *Journal of Financial Services Research*, 47(1), 135-152.
- le Calvez, A., & Cliff, D. (2018, November). Deep learning can replicate adaptive traders in a limit-order-book financial market. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1876-1883). IEEE.
- Lechman, E., & Marszk, A. (2017). Information and Communication Technologies for Economic Development. In *Catalyzing Development through ICT Adoption* (pp. 3-14). Springer, Cham.
- Lima, PA & Sichman, JS. (2007). Towards Cooperation Among Competitive Trader Agents. *ICEIS*, (4), 138-143.
- Litzenberger, R., Castura, J & Gorelick, R. (2012). The impacts of automation and high frequency trading on market quality. *Annual Review of Financial Economics*, 4(1), 59-98.
- Lozada, J. (2014). Investigación Aplicada: Definición, propiedad intelectual e industria. *CienciaAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica*, 3(1), 47-50.
- Ma, C., Zhang, J., Liu, J., Ji, L., & Gao, F. (2021). A parallel multi-module deep reinforcement learning algorithm for stock trading. *Neurocomputing*, 449, 290-302.
- Mankad, S., Michailidis, G & Kirilenko, A (2013). Discovering the ecosystem of an electronic financial market with a dynamic machine-learning method. *Algorithmic Finance*, 2(2), 151-165.
- Maringer, D., & Ramtohul, T. (2010, April). Threshold recurrent reinforcement learning model for automated trading. In *European Conference on the Applications of Evolutionary Computation* (pp. 212-221). Springer, Berlin, Heidelberg.
- Masteika, S., & Vaitonis, M. (2015, June). Quantitative research in high frequency trading

- for natural gas futures market. In *International Conference on Business Information Systems* (pp. 29-35). Springer, Cham.
- Medina, L. Á. (2003). Aplicación de la teoría del portafolio en el mercado accionario colombiano. *Cuadernos de economía*, 22(39), 129-168.
- Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of financial Markets*, 16(4), 712-740.
- Mirabovich, T. (2014). Academic citation indices and databases. *Law and modern states*, (3).
- Morales, J. J. R., Yopez, E. R., & Castañeda, B. G. (2015). Hipótesis de Mercados Eficientes en el Mercado Accionario Colombiano a través del Estudio de Eventos. In *Memorias XX Congreso Internacional de Contaduría, Administración e Informática*.
- Nettles, J., Brayer, N., Jenner, C., Ngo, A., Putnam, C., Shank, T., ... & Beling, P. (2015, April). Forecasting intraday volume distributions. In *Systems and Information Engineering Design Symposium (SIEDS), 2015* (pp. 97-102). IEEE.
- Ng, C. Y., Mustafa, Z., & Kurian, J. (2019). Impact of International Co-Authorships to a Young Malaysian University Specialising in Science, Technology, Engineering and Mathematics. *DESIDOC Journal of Library & Information Technology*, 39(5), 238-243.
- Panourgias, N.S. (2015). Capital markets integration: A sociotechnical study of the development of a cross-border securities settlement system. *Technological Forecasting & Social Change*, 99, 317-338.
- Parkes, D. C., & Wellman, M. P. (2015). Economic reasoning and artificial intelligence. *Science*, 349(6245), 267-272.
- Pellens, M., da Silva, M & da Silva, AT. (2017). Augmented Reality and Serious Games: A Systematic Literature Mapping. In *19th Symposium on Virtual and Augmented Reality (SVR)* (pp. 227-235), IEEE.
- Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12* (pp. 1-10).
- Pino, A., Uribe, J. M., & Jiménez, D. M. (2018). Relevancia de los inversionistas institucionales en el mercado accionario colombiano. *Semestre Económico*, 20(44), 45-65.
- Rankia (2017). *¿Qué es stop loss y take profit?*. <https://www.rankia.co/blog/analisis-colcap/3601459-que-stop-loss-take-profit>

- Rankia (2020). *Los 10 mejores ETFs: NASDAQ-100, S&P 500, Oro, Tecnologías de la información y Biotecnología*. <https://www.rankia.com/blog/opiniones/4585266-10-mejores-etfs-nasdaq-100-s-p-500-oro-tecnologias-informacion-biotecnologia>
- Raudys, A., Plikynas, D & Masteika, S. (2014). Novel Automated Multi-Agent Investment System Based on Simulation of Self-Excitatory Oscillations. *Transformation in Business & Economics*, 13(2).
- Roledene, S., Ariyathilaka, L., Liyanage, N., Lakmal, P., & Bamunusinghe, J. (2016, December). GeniBux-event based intelligent Forex trading strategy enhancer. In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)* (pp. 1-6). IEEE.
- Ruta, A., Ruta, D & Cen, L. (2017). Algorithmic Daily Trading Based on Experts' Recommendations. *International Symposium on Methodologies for Intelligent Systems*, Springer, 735-744.
- Seo, J. Y., & Chai, S. (2013). The role of algorithmic trading systems on stock market efficiency. *Information Systems Frontiers*, 15(5), 873-888.
- Sezer, O. B., Ozbayoglu, A. M., & Dogdu, E. (2017, April). An artificial neural network-based stock trading system using technical analysis and big data framework. In *Proceedings of the South East Conference* (pp. 223-226). ACM.
- Ta, V. D., Liu, C. M., & Addis, D. (2018, December). Prediction and portfolio optimization in quantitative trading using machine learning techniques. In *Proceedings of the Ninth International Symposium on Information and Communication Technology* (pp. 98-105).
- Urbizagástegui-Alvarado, R., & Cortés-Contreras, M. T. (2002). La productividad de autores en la Revista Geológica de Chile. *Ciencias de la Información*, 33(2), 15-25.
- Van Kervel, V., & Menkveld, A. J. (2019). High-frequency trading around large institutional orders. *The Journal of Finance*, 74(3), 1091-1137.
- Vella, V., & Ng, W. L. (2014). Enhancing risk-adjusted performance of stock market intraday trading with neuro-fuzzy systems. *Neurocomputing*, 141, 170-187.
- Veryzhenko, I., Arena, L., Harb, E., & Oriol, N. (2017). Time to slow down for high-frequency trading? Lessons from artificial markets. *Intelligent Systems in Accounting, Finance and Management*, 24(2-3), 73-79.
- Villada, F., Muñoz, N., & García, E. (2012). Aplicación de las Redes Neuronales al Pronóstico de Precios en el Mercado de Valores. *Información tecnológica*, 23(4), 11-20.

Zhang, H., Li, Z., Shahriar, H., Tao, L., Bhattacharya, P., & Qian, Y. (2019, July). Improving prediction accuracy for logistic regression on imbalanced datasets. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 918-919). IEEE.