

Escuela Superior Politécnica del Litoral

Facultad de Ciencias Sociales y Humanísticas

Análisis del Comportamiento de Clientes del Mercado de Valores Mediante
Machine Learning e Inteligencia Artificial
ADMI-1254

Proyecto Integrador

Previo la obtención del Título de:

Economista

Presentado por:

Hugo David Gaibor Sarmiento

Giovanna Paulette Segarra Fernández

Guayaquil - Ecuador

Año: 2025

DEDICATORIA

El presente proyecto lo dedico principalmente a mis padres, quienes me brindaron su apoyo a lo largo de mi carrera universitaria, siempre dispuestos a ayudarme cuando lo necesitaba. También a Dios, mis hermanas, quienes con sus consejos y guía han sido fundamentales en mi vida profesional.

Giovanna Paulette Segarra Fernandez

A Dios, a mis padres por brindarme todo su apoyo desde el comienzo de mi etapa en la vida universitaria, a pesar de las adversidades del camino y a mis hermanos por su apoyo incondicional y consejos en mi vida académica y profesional.

Hugo David Gaibor Sarmiento

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a mi compañero de tesis y mejor amigo, David Gaibor, por acompañarme y brindarme su apoyo incondicional a lo largo de casi toda mi etapa universitaria. Extiendo también mi gratitud a mis amigos Johnny Avilez, Manuel Naranjo y Andrés Asanza, quienes con sus ocurrencias, ánimos y palabras llenas de cariño siempre lograron sacarme una sonrisa. Sé que me faltan muchos más amigos por nombrar, pero cada uno ha dejado una huella especial en este camino. A mi familia, gracias por celebrar cada uno de mis logros con alegría y orgullo, convirtiéndose en mi mayor fuente de motivación. Finalmente, a mi tutor Bernard, a quien siempre recordaré con aprecio. Sus consejos, enseñanzas y experiencias compartidas en nuestras reuniones fueron inspiración constante y nos impulsaron a alcanzar nuestras metas.

Giovanna Paulette Segarra Fernández

Quiero expresar mi más profundo
agradecimiento a mis padres y hermanos
quienes han sido mi inspiración y mis pilares
fundamentales, a mi compañera de tesis y
mejor amiga, Giovanna Segarra por haberme
brindado su apoyo, compañía y consejos
durante toda mi etapa universitaria. Ah,
Johnny Avilés y Pablo Zúñiga, las risas y su
motivación han hecho de este proceso algo
más llevadero y memorable. Dedico este
logro a mi novia Romina, le agradezco por
recordarme la importancia de disfrutar cada
momento, su alegría y apoyo han sido
esenciales en esta etapa de culminación de la
carrera. A su vez, extendiendo mis
agradecimientos a todas aquellas personas
que han estado presentes durante este
proceso, cada uno ha dejado una huella en mi
vida y carrera profesional, gracias por ser
parte de mi historia. A todos ustedes, con
mucho amor, cariño y orgullo, les dedico este
logro, los llevo conmigo siempre.

Hugo David Gaibor Sarmiento

DECLARACIÓN EXPRESA


Nosotros Giovanna Paulette Segarra Fernandez y Hugo David Gaibor Sarmiento acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me/nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi/nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique los autores que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 30 de mayo del 2025.


Giovanna Paulette
Segarra Fernández


Hugo David Gaibor
Sarmiento

EVALUADORES

María Cristina Aguirre Valverde

Profesor de Materia

Henry Bernard Moscoso Miranda

Tutor de proyecto

RESUMEN

En el mercado de valores ecuatoriano, las casas de valores enfrentan el reto de perfilar sus estrategias comerciales al tipo de cliente segmentado con base en información respaldada por sus negociaciones históricas. Sin embargo, las propuestas de inversión se apoyan principalmente en la experiencia del operador o asesor comercial, lo que introduce un alto grado de subjetividad. Una de las principales limitaciones es la ausencia de herramientas avanzadas capaces de transformar grandes volúmenes de datos en conocimiento útil, lo que restringe el diseño de estrategias efectivas y disminuye la competitividad de estas instituciones. El presente proyecto tiene como objetivo analizar el comportamiento de los clientes mediante técnicas de Machine Learning, integrando variables tradicionales para segmentar e identificar perfiles de clientes. Para ello, se utilizaron datos históricos del período 2014–2024 y se aplicaron modelos supervisados (Random Forest) y no supervisados (K-Means), complementados con validación externa a través de asesores financieros. Los resultados evidencian que factores como el monto a invertir, el valor efectivo y el total de la operación son más determinantes que variables técnicas convencionales, permitiendo una segmentación más precisa y el diseño de estrategias comerciales diferenciadas. En conclusión, el proyecto propone una herramienta innovadora que optimiza la gestión de clientes en casas de valores, favorece la toma de decisiones estratégicas y fortalece la posición competitiva en el mercado bursátil.

Palabras clave: Mercado de Valores, Machine Learning, Segmentación de Clientes.

ABSTRACT

In the Ecuadorian stock market, brokerage firms face the challenge of tailoring their business strategies to specific customer segments based on information supported by their historical trading records. However, investment proposals are mainly based on the experience of the trader or commercial advisor, which introduces a high degree of subjectivity. One of the main limitations is the absence of advanced tools capable of transforming large volumes of data into useful knowledge, which restricts the design of effective strategies and reduces the competitiveness of these institutions. This project aims to analyze customer behavior using machine learning techniques, integrating traditional variables to segment and identify customer profiles. To this end, historical data from the period 2014–2024 was used and supervised (Random Forest) and unsupervised (K-Means) models were applied, complemented by external validation through financial advisors. The results show that factors such as the amount to be invested, the effective value, and the total value of the transaction are more decisive than conventional technical variables, allowing for more accurate segmentation and the design of differentiated commercial strategies. In conclusion, the project proposes an innovative tool that optimizes customer management in securities firms, promotes strategic decision-making, and strengthens competitive position in the stock market.

Keywords: *Stock Market, Machine Learning, Customer Segmentation.*

ÍNDICE GENERAL

Resumen.....	I
Abstract.....	II
Índice General.....	III
Índice de Figura	V
Índice de Tablas	V
Capítulo 1.....	1
1. Introducción	2
1.1 Descripción del problema.....	4
1.2 Justificación del problema.....	4
1.3 Objetivos.....	5
1.3.1 Objetivo general.....	5
1.3.2 Objetivos específicos	5
1.4 Marco teórico - necesidades de mercado de valores, uso de ML e IA.....	6
Capítulo 2.....	14
2. Metodología	15
2.1 Fuentes de Datos:	15
2.2 Limpieza y análisis exploratorio de datos	18
2.3 Procesamiento de datos	18
2.3 Modelos aplicados	19
2.3.1 Modelo supervisado: Random Forest	19
2.3.2 Modelo no supervisado: K-means con MCA (Análisis de Correspondencia Múltiples)	22
2.5 Validación externa	23
2.6 Justificación del enfoque metodológico	24

Capítulo 3.....	25
3. Resultados.....	26
3.1 Caracterización general del conjunto de datos:.....	26
3.1.1 Descripción del dataset.....	26
3.1.2 Análisis de valores nulos y tratamiento aplicado	26
3.2 Resultados del análisis exploratorio de datos (EDA)	27
3.2.1 Detección y tratamiento de valores atípicos	27
3.3 Resultados del modelo supervisado: Random Forest	28
3.3.1 Variables seleccionadas tras análisis de importancia	28
3.3.2 Rangos de clase	29
3.3.3 Evaluación del Desempeño del Modelo.....	31
3.3.4 Interpretación de resultados: factores que predicen el volumen a negociar del cliente	33
3.4 Resultados del modelo no supervisado: k-means	35
3.4.1 Selección del número óptimo de clústers	35
3.4.2 Caracterización de los clústers: perfil del cliente por grupo	36
3.4.3 Interpretación de los grupos: diferencias clave entre perfiles	38
3.4.4 Métricas de Evaluación	44
Capítulo 4.....	47
4. Conclusiones y recomendaciones	48
4.1 Conclusiones	48
4.2 Recomendaciones	49

ÍNDICE DE FIGURA

Figura 1	28
Figura 2	30
Figura 3	32
Figura 4	35
Figura 5	36
Figura 6	37
Figura 7	40

ÍNDICE DE TABLAS

Tabla 1	27
Tabla 2	29
Tabla 3	31
Tabla 4	34
Tabla 5	38
Tabla 6	41
Tabla 7	44

CAPÍTULO 1

1. INTRODUCCIÓN

En la dinámica actual del mercado de valores ecuatoriano, las casas de valores presentan los desafíos crecientes por adaptarse a las nuevas exigencias del entorno financiero, que se caracterizan por la evolución constante de productos de inversión, diversificación de perfiles de riesgo y el acceso progresivo de los distintos segmentos de la población hacia los servicios bursátiles. Como mencionan Laurisch et al. (2024), es crucial que las empresas comprendan el comportamiento del cliente, ya que influye en los productos, los precios y la estrategia de fidelización, al realizar un análisis del comportamiento de los clientes, las empresas pueden adaptar estrategias para convertir a su público objetivo en compradores y segmentarlos. Una de las limitaciones significativas de las instituciones en los mercados de valores es el de trabajar con grandes volúmenes de datos históricos sobre transacciones, activos financieros e información de clientes en conocimiento útil para la toma de decisiones estratégicas y colocación de ofertas en el mercado.

Tradicionalmente, el análisis del comportamiento de los clientes y su respectiva clasificación se ha basado en experiencia previa de asesores comerciales, enfoques manuales de clasificación o métodos estadísticos convencionales, mismos que, a medida que el mercado actual se desarrolla se vuelven inefficientes. Frente a este panorama, se hace evidente la necesidad de adoptar herramientas avanzadas que permitan un análisis automatizado y predictivo del comportamiento de los inversionistas en el mercado de valores. Para obtener una ventaja competitiva, la cantidad de datos junto con los algoritmos que encuentren los productos que mejor se adapten a las necesidades de los clientes serán clave (Francés Monedero, 2020).

La inteligencia artificial (IA) y el machine learning (ML) surgen como tecnologías clave para resolver esta problemática, como menciona Alonso Robisco & Carbó (2022), los modelos tradicionales econométricos usados en los servicios financieros y en la economía pueden

complementarse y, en ciertos casos, sustituirse con técnicas provenientes de la IA. Estas tecnologías en conjunto ofrecen un conjunto de métodos y modelos capaces de poder identificar patrones complejos en grandes volúmenes de datos, segmentar a clientes de manera precisa y predecir comportamientos futuros, apoyando de esta manera la toma de decisiones con un nivel alto de adaptabilidad y precisión ante cambios actuales. Al analizar las emociones, actitudes, preferencias y motivaciones de los clientes, las empresas pueden adaptar estrategias para convertir a su público objetivo en compradores y segmentarlos según sus preferencias para impulsar una mayor interacción (Laurisch et al., 2024). Esto permite que las casas de valores analicen de mejor manera a sus clientes, desarrollen estrategias comerciales efectivas, optimicen recursos y procesos, obteniendo una mayor capacidad de innovación en el diseño y promoción de sus productos financieros.

El objetivo del presente proyecto integrador es analizar el comportamiento de los clientes del mercado de valores ecuatoriano mediante técnicas avanzadas de Inteligencia Artificial y Machine Learning, integrando variables tradicionales y no tradicionales, para identificar y segmentar clientes potenciales en función de las ofertas disponibles en la casa de valores. Para cumplir con este objetivo se plantea identificar a los clientes potenciales a partir del análisis de su historial de inversiones, diseñar un prototipo funcional que integre técnicas de IA y Machine Learning para optimizar la clasificación y el análisis del comportamiento de los inversionistas y del mismo modo, se plantea proponer recomendaciones estratégicas para mejorar la colocación de ofertas en el mercado de valores.

1.1 DESCRIPCIÓN DEL PROBLEMA

En el contexto actual del mercado de valores ecuatoriano, las casas de valores que actúan como intermediarios bursátiles, enfrentan importantes desafíos para comprender y anticipar el comportamiento de sus usuarios, pese a disponer de amplios datos históricos de transacciones, perfiles y activos. La falta de herramientas avanzadas para procesar y analizar información limita el diseño de estrategias comerciales personalizadas, la identificación de segmentos de alto potencial y la proyección precisa del volumen transaccional.

Los enfoques tradicionales, basados en experiencia o análisis manual, resultan insuficientes debido a la creciente complejidad del mercado, caracterizada por una mayor diversificación de productos y perfiles de riesgo. Por ende, esta situación reduce la capacidad de fidelización de clientes, el control sobre los flujos operativos y la posición competitiva que podría tener una casa de valores en el mercado (Sydle, s.f.). Ante esto, se vuelve necesario aplicar modelos integradores de aprendizaje automático que permitan segmentar, predecir preferencias y estimar volúmenes, optimizando así la toma de decisiones estratégicas dentro de la casa de valores.

1.2 JUSTIFICACIÓN DEL PROBLEMA

Las casas de valores cuentan con tres principales fuentes de ingresos: administración de portafolios, estructuración financiera y corretaje de valores. Sin embargo, dos de estas dependen en gran medida del corretaje de valores, también conocido como trading a nivel internacional. Este servicio consiste en la realización de negociaciones a través de los sistemas bursátiles provistos por la Bolsa de Quito y la Bolsa de Guayaquil, en los cuales se transan diversos productos financieros disponibles en el mercado.

Estos sistemas recolectan datos relevantes de cada cliente que negocia a través de la casa de valores. No obstante, dicha información no ha sido procesada ni aprovechada de manera sistemática. Actualmente, los agentes comerciales se basan principalmente en su experiencia personal y en el conocimiento empírico que tienen sobre los clientes asignados, lo que genera procesos altamente individualizados, poco estandarizados y propensos a errores. Esta dependencia de la memoria y percepción subjetiva impide una atención más eficiente, coherente y sostenible en el tiempo.

Por ello, se vuelve fundamental implementar un modelo que centralice, organice y analice los datos históricos y actuales de los clientes. Esto permitiría una mejor comprensión de su comportamiento, la optimización de estrategias comerciales y la toma de decisiones basada en evidencia, lo cual es esencial para fortalecer la estabilidad y el crecimiento del área de corretaje y, por ende, del negocio en general.

1.3 OBJETIVOS

1.3.1 Objetivo general

Analizar el comportamiento de clientes del mercado de valores mediante técnicas avanzadas de Machine Learning integrando variables de fuentes externas para la identificación y segmentación de clientes potenciales según las ofertas disponibles en la casa de valores.

1.3.2 Objetivos específicos

- Identificar a los clientes potenciales a partir del análisis de su historial de inversiones, con el propósito de ofrecerles productos financieros personalizados que se ajusten a sus preferencias previas.

- Determinar un conjunto de variables disponibles de los clientes y de fuentes externas que permitan innovar en la identificación de clientes potenciales.
- Diseñar un modelo que integre técnicas de inteligencia artificial y aprendizaje automático para optimizar la clasificación y análisis del comportamiento de los clientes de las casas de valores.
- Proponer recomendaciones de los factores que determinan la colocación de ofertas comerciales en base a segmentaciones de clientes.

1.4 MARCO TEÓRICO - NECESIDADES DE MERCADO DE VALORES, USO DE ML E IA

La historia del mercado de valores en Ecuador se remonta al Código de Comercio de 1882, donde ya se hace mención a las bolsas de comercio. No obstante, fue el Código de 1906, durante el gobierno de Eloy Alfaro, el que estableció un marco más estructurado para las actividades mercantiles. En 1935 se constituyó la Bolsa de Valores y Productos del Ecuador C.A., en la ciudad de Guayaquil, aunque su funcionamiento fue efímero, operando por apenas tres años. Según la Superintendencia de Compañías, Valores y Seguros (s.f.), el limitado desarrollo del mercado en esa época se debió principalmente a la escasa educación financiera, a la débil estructura industrial del país y a la inestabilidad política.

En 1964, se creó la Comisión de Valores dentro de la Corporación Financiera Nacional, como una entidad promotora del crédito y del desarrollo industrial. Uno de sus mandatos legales era fomentar la constitución de bolsas de valores. Fue así que, en 1969, con el respaldo del crecimiento económico y mediante disposición legal de la Comisión Legislativa Permanente, se autorizó la creación de las bolsas de valores en Quito y Guayaquil, organizadas como compañías anónimas conforme a la Ley No.111 Su función principal consistía en ofrecer mecanismos para

la negociación de valores en condiciones de seguridad, equidad, transparencia y con precios justos.

Finalmente, el 28 de mayo de 1993 se expidió la primera Ley de Mercado de Valores, que estableció a las casas de valores como los únicos intermediarios autorizados para operar en el mercado bursátil, gestionando la compraventa de valores por cuenta de sus comitentes (Ley de Mercado de Valores, 1993). Estas entidades, constituidas como compañías anónimas, deben contar con autorización y supervisión de la Superintendencia de Compañías, Valores y Seguros, así como, con su inscripción en el Catastro Público del Mercado de Valores para operar en sistemas bursátiles autorizados.

Además de su rol como intermediarias en el mercado bursátil, las casas de valores pueden desarrollar actividades propias de la banca de inversión, orientadas a identificar y estructurar opciones de financiamiento e inversión para sus clientes. Estos servicios están dirigidos tanto a personas naturales como jurídicas, conocidas como comitentes, quienes delegan a las casas de valores funciones como la estructuración de emisiones, la asesoría financiera o la gestión de portafolios mediante mandatos u órdenes de negociación. Estas operaciones son ejecutadas por operadores de valores, quienes actúan en representación de la casa de valores y deben estar debidamente calificados e inscritos en el Catastro Público del Mercado de Valores.

En el mercado de valores ecuatoriano, los clientes pueden clasificarse en dos grandes grupos: clientes institucionales y clientes naturales. Los primeros comprenden entidades tanto públicas como privadas, tales como fondos de inversión, aseguradoras, bancos, cooperativas financieras y empresas del sector real (Primicias, s.f.). Estas instituciones pueden desempeñar un doble rol: como emisores, al colocar instrumentos financieros en el mercado para captar recursos, y como inversionistas, al destinar sus excedentes de liquidez a la adquisición de valores

negociables. En cambio, los clientes naturales, en su mayoría personas individuales, suelen participar únicamente como inversionistas, guiando sus decisiones de inversión conforme a su perfil de riesgo y a sus objetivos financieros.

Dentro de la estructura del mercado, se encuentran instituciones complementarias que garantizan su funcionamiento ordenado y transparente. Entre ellas se destacan las calificadoras de riesgo, que emiten opiniones técnicas sobre la solvencia y capacidad de pago de los emisores, permitiendo a los inversionistas evaluar el nivel de riesgo asociado a cada instrumento financiero (Superintendencia de Compañías, Valores y Seguros, s.f.). De igual forma, operan los depósitos centralizados de valores, como Decevale (DCV privado) y el Depósito Centralizado de Valores del Banco Central del Ecuador (DCV público), cuya función principal es la custodia, transferencia y liquidación de los valores negociados en el mercado bursátil (Decevale, s.f.).

Los productos financieros negociables, en el mercado de valores ecuatoriano, se agrupan principalmente en función del tipo de mercado, del instrumento y del emisor. Según el momento de la colocación, se distingue entre el mercado primario, donde los valores son emitidos por primera vez, y el mercado secundario, donde estos valores se compran y venden entre inversionistas.

Los títulos que se negocian pueden clasificarse en dos grandes grupos: instrumentos de renta fija, como los bonos, que ofrecen rendimientos preestablecidos; e instrumentos de renta variable, como las acciones, cuyo retorno depende del desempeño de la empresa emisora. A su vez, estos instrumentos pueden provenir de emisores del sector público (como el Estado o empresas públicas) o del sector privado (como compañías privadas que buscan financiamiento a través del mercado).

Actualmente, el mercado de valores ecuatoriano presenta una diversidad de instrumentos: bonos corporativos, obligaciones del Estado, papeles comerciales, certificados de inversión, entre otros. Esta variedad representa oportunidades, pero también retos complejos en la evaluación y comparación de alternativas, dado el dinamismo del mercado (MINTEL, 2020).

La dinámica del mercado bursátil moderno exige una capacidad de análisis que permita considerar múltiples variables financieras, económicas y temporales en tiempo real. Sin embargo, el análisis tradicional, basado en modelos financieros estáticos y procesos manuales, presenta limitaciones relevantes frente a la complejidad de los datos actuales. Estas metodologías, si bien útiles, suelen ser insuficientes para capturar patrones ocultos o prever comportamientos futuros del mercado con precisión (Dútan Pugo, 2021).

Ante esta situación, surge la necesidad de incorporar tecnologías avanzadas como el aprendizaje automático (machine learning), la inteligencia artificial y el análisis automatizado de datos. Estas herramientas permiten procesar grandes volúmenes de datos, identificar tendencias emergentes y ofrecer recomendaciones de inversión más informadas y personalizadas (Proyecto X-pedientes Económicos, 2024). En Ecuador específicamente, ya existen estudios que aplican redes neuronales para optimizar portafolios en la Bolsa de Valores de Quito (Ochoa Luzuriaga, 2024), y también se han realizado diagnósticos sobre la adopción de IA en el sector financiero (MINTEL, 2020).

La incorporación del Machine Learning y la Inteligencia Artificial ha tenido una tendencia creciente en los últimos años debido al gran impacto de su aplicación en distintas áreas, destacándose en el área financiera, en donde ha tenido un gran alcance en la clasificación y segmentación de clientes. Los resultados de la aplicación de modelos de ML, enfatizan la eficacia en los resultados de las redes neuronales artificiales (RNA) y los modelos de conjunto,

como los bosques aleatorios, para pronosticar con precisión las empresas en dificultades financieras que cotizan en TSX (Lokanan ME y Ramzan S, 2024).

En un entorno en donde los datos presentan un crecimiento exponencial con una gran variedad de perfiles de inversores, los enfoques clásicos de resultan ser poco precisos y limitados, por otra parte, los métodos de ML permiten procesar grandes volúmenes de información revelando patrones complejos, además que sus modelos se adaptan conforme se ingresa nueva información, generando una mayor precisión y alcance de sus modelos. De acuerdo con la investigación de González et al. (González et al., 2021) el Big Data y sus aplicaciones son esenciales para el análisis en tiempo real, permitiendo a las operadoras actuar de manera eficiente en un mercado que cambia rápidamente. Según (Sufian Mozumder et al., 2024), la aplicación de algoritmos de Machine Learning en la segmentación de clientes ha demostrado ser muy beneficiosa para la banca, permitiéndoles comprender de mejor manera a sus clientes y adaptar los servicios de manera más efectiva (p. 06).

Las aplicaciones de la IA y el ML en la actualidad son numerosas, desde su aplicación en áreas de la salud, construcción, educación, finanzas, etc (Fuentes et al., 2019; Castro et al., 2020; Castaneda, 2023; Verduga et al., 2024) . El sistema financiero fue una de las primeras áreas en las que se aplicó estas herramientas, por ejemplo en mercados como el de Wall Street, el software IBM Watson es de vital importancia, ya que realiza la mayoría de las operaciones (Denniye, 2018). A su vez, en el sistema financiero ecuatoriano, instituciones como el SRI han desarrollado modelos inteligentes, estos modelos permiten realizar una selección objetiva de los contribuyentes que se encuentren relacionados con tramas de evasión y elusión tributaria, cuyos resultados permitirán mejorar a sus modelos paulatinamente a medida que actualizan los datos (SRI, 2018).

La aplicación de tecnologías avanzadas, como las Redes Neuronales avanzadas, no solo mejora la toma de decisiones internas, sino también puede atraer a inversores interesados en participar en un mercado respaldado por análisis avanzados y estrategias de inversión innovadoras (Ochoa, 2024). Las técnicas de aprendizaje automático son empleadas para identificar patrones en los movimientos del mercado y predecir futuras tendencias de precios, lo que ayuda a los gestores de cartera a tomar decisiones más informadas y rápidas (González et al., 2021).

Los algoritmos de aprendizaje automático pueden identificar actividades inusuales con información inmediata, lo cual permite a las casas de valores reaccionar prontamente ante sospechas de fraude y riesgo. Al utilizar algoritmos para analizar datos económicos y financieros, las casas de valores pueden prever situaciones adversas antes de que ocurran, implementando medidas preventivas para mitigar los efectos negativos (Figuerola et al., 2021). A través del análisis de datos sobre el comportamiento y preferencias de los clientes, las casas de valores pueden ofrecer servicios financieros más adaptados a las necesidades individuales de los inversores. Las plataformas de inversión globales, han comenzado a implementar chatbots impulsados por IA que puedan interactuar con los usuarios, proporcionar información sobre el mercado y ofrecer recomendaciones basadas en algoritmos avanzados (Guerrero, 2023).

Para efectos de la presente tesis, se consideran las siguientes definiciones técnicas que se van a utilizar. El machine learning reconoce patrones, extrae conocimiento y genera predicciones, mientras más datos tenga mejores resultados va a generar, algunas de las técnicas usadas para generar más datos son:

Análisis de regresión: Es un método de aprendizaje supervisado, en el cual se busca establecer una relación entre dos variables diferentes, estimando de forma efectiva como una variable afectará a la otra.

Clasificación: Es un método de aprendizaje supervisado, se busca categorizar un conjunto de datos en diferentes clases, el modelo se entrena con datos etiquetados según diferentes clases, algunas de las aplicaciones comunes incluyen la detección facial, reconocimiento de escritura a mano, entre otras.

Clustering: Es un método de aprendizaje no supervisado, el cual trabaja en descubrir automáticamente patrones naturales o grupos que se producen dentro de un conjunto de datos, algunas de sus aplicaciones son: segmentación de mercados, filtrado de spam por correo, entre otros.

Árboles de decisión: Es un método de aprendizaje supervisado, que categoriza un grupo de objetos en función de cuestiones relativas a sus cualidades en los puntos nodales correspondientes, algunas de sus aplicaciones incluyen la detección de fraudes, gestión de relaciones con clientes, marketing, predicción de precios, entre otras.

La propuesta de valor en el presente proyecto se basa en crear un prototipo funcional de software que integre Machine learning e Inteligencia Artificial, que permita una optimización de procesos en el análisis del comportamiento de los clientes en las casas de valores, además de poder identificar qué características hacen a un cliente ser potencial en las casas de valores, con el propósito de ofrecer productos financieros que se adapten al perfil de estos.

Se espera tener un impacto comercial significativo, ya que al poder segmentar clientes de casas de valores en base información obtenida de fuentes externas de los mismos e información sintética, desarrollando con ello un modelo integrador y adaptable a las distintas áreas del

sistema financiero ecuatoriano en cuanto a la identificación de clientes potenciales, análisis del comportamiento de cliente y la clasificación de estos, optimizando de esta manera los recursos de las organizaciones, mejorando su enfoque en cuanto a los servicios que ofrecen a los clientes y fortaleciendo la toma de decisiones en cuanto a la estabilidad y crecimiento de la organización.

CAPÍTULO 2

2. METODOLOGÍA

2.1 FUENTES DE DATOS:

Los datos utilizados en el presente proyecto provienen de fuentes públicas de diversos actores del mercado de valores, ajustándose al contexto de la realidad ecuatoriana, con un enfoque particular en el mercado del litoral. La base de datos se compone principalmente de registros históricos de negociación correspondientes al período 2014–2024, complementados con información simulada para enriquecer el análisis. El conjunto de datos consta de un total de 28,087 observaciones e incluye tanto variables numéricas como categóricas

Para identificar patrones en el comportamiento de los clientes, se adoptó un enfoque metodológico cuantitativo, en virtud de la necesidad de analizar grandes volúmenes de datos estructurados. Se aplicó el modelo de aprendizaje supervisado Random Forest, con el objetivo de segmentar clientes potenciales en función de su comportamiento histórico, optimizando así la personalización de ofertas comerciales y atendiendo de manera más eficiente sus necesidades. Complementariamente, se utilizó el modelo de aprendizaje no supervisado K-means, orientado a descubrir patrones emergentes y agrupar a los clientes según características como comportamiento de inversión, perfil de riesgo, nivel de ingresos, entre otros.

Para el desarrollo del modelo de segmentación y predicción se utilizarán variables operativas, financieras y del perfil de los clientes. A continuación, se describen las principales variables que forman parte del dataset, junto con su significado y relevancia para el análisis de comportamiento de los clientes.

- **Hora Orden:** Hora exacta en la que se ingresa la orden de operación en el sistema.
- **Hora Cierre:** Hora en la que se cierra la operación bursátil.

- **Fecha de operación:** Fecha en la que se crea la orden de negociación en el sistema.
- **Fecha valor:** Fecha en la que una negociación se considera efectivamente realizada
- **Fecha de emisión:** Fecha en que se emitió el título negociado.
- **Tipo de operación:** Naturaleza de la operación: compra o venta.
- **Emisor:** Nombre de la entidad emisora del título negociado.
- **Título:** Tipo de instrumento negociado (acciones, bonos, etc)
- **Plazo Remanente:** Tiempo restante para el vencimiento del título (si aplica).
- **Tasa de interés:** Tasa de interés pactada para el instrumento (si aplica).
- **Cantidad / Valor nominal:** Número de títulos o valor nominal involucrado en la operación.
- **Precio Cierre:** Precio al que se cierra la transacción en el mercado.
- **Valor efectivo:** Monto efectivo involucrado en la operación.
- **Interés Transcurrido:** Interés acumulado hasta la fecha de operación (si aplica).
- **Rendimiento:** Rentabilidad estimada de la inversión.
- **% Comisión Bolsa de Valor:** Porcentaje de comisión cobrada por la Bolsa de Valores.
- **Comisiones Bolsa de Valores:** Valor monetario de la comisión de la Bolsa.
- **% Comisión Casa de Valor:** Porcentaje de comisión cobrada por la Casa de Valores.
- **Comisiones Casa de Valores:** Valor monetario de la comisión de la Casa de Valores.

- **Total Comisión:** Total de comisiones generadas en la operación.
- **% Comisión total:** Porcentaje total de comisión sobre el valor de la operación.
- **Valor total:** Valor total de la operación incluyendo comisiones e impuestos.
- **Cupón:** Monto del pago periódico de interés (si aplica).
- **Mercado:** Tipo de mercado en el que se realiza la operación (mercado bursátil, extrabursátil, etc.).
- **Sucursal:** Oficina de la Casa de Valores donde se registra la operación.
- **Renta:** Tipo de renta del instrumento: fija o variable.
- **Fecha de vencimiento:** Fecha en que vence el instrumento negociado.
- **Monto a invertir:** Monto invertido.
- **No. Emisión:** Número o código de emisión del título.
- **Tipo Corretaje:** Tipo de intermediación realizada: nuevo producto, corretaje comercial, etc.
- **Sector:** Sector económico del emisor (financiero, industrial, etc).
- **Segmento de Mercado:** Indicador si la operación pertenece a un mercado secundario o primario.
- **Retención Bolsa:** Valor de retención aplicada por la Bolsa.
- **Retención Casa:** Valor de retención aplicada por la Casa de valores.
- **Cámara:** Cámara de compensación involucrada en la operación (Decevale privado o Decevale público)
- **Programa:** Número de emisión respecto al producto negociado en el mercado.
- **ID anónimo del cliente:** Codificación personalizada para resguardar la identidad.
- **Tipo de persona:** Natural o jurídica.

- **Ciudad de residencia.**
- **Ciudad operación:** Ciudad donde se crea la operación.
- **Perfil de riesgo (Categoría):** Moderado, otras opciones.
- **Precio:** Porcentaje pagado o recibido sobre el valor del instrumento.
- **Plazo del instrumento financiero.**
- **Calificación de riesgo:** Emitida por agencias autorizadas.
- **Calificadora:** Empresa calificadora que asigna la calificación de riesgo.
- **Fecha de calificación:** Fecha en la que se emitió la calificación de riesgo.

Estas variables fueron almacenadas en archivos estructurados (formatos Excel o CSV), los cuales fueron posteriormente depurados y organizados mediante Power Query para su análisis. No se incorporó directamente la variable “frecuencia de operación” debido a la falta de estandarización en su registro, aunque se consideró su análisis desde un enfoque agregado. La variable dependiente del estudio fue perfil de riesgo.

2.2 LIMPIEZA Y ANÁLISIS EXPLORATORIO DE DATOS

Se ejecutó un proceso de limpieza que incluyó la eliminación de valores nulos, inconsistencias y registros duplicados. Posteriormente, se realizó un análisis exploratorio de datos (EDA), cuyo objetivo fue comprender la distribución de las variables, detectar valores atípicos, identificar correlaciones preliminares y establecer transformaciones necesarias para el modelado posterior.

2.3 PROCESAMIENTO DE DATOS

El tratamiento de los datos se llevó a cabo en el lenguaje de programación Python, utilizando principalmente las librerías NumPy y Pandas. El conjunto de datos fue dividido en un

80% para entrenamiento y 20% para validación. Asimismo, se contempló la incorporación futura de técnicas como web scraping para integrar datos externos relevantes, tales como el cumplimiento tributario ante el SRI y el pago de impuestos a la renta, con el fin de enriquecer las predicciones y robustecer los modelos generados.

2.3 MODELOS APLICADOS

2.3.1 Modelo supervisado: Random Forest

Se empleó el algoritmo Random Forest, el cual consistió en el ensamble de múltiples árboles de decisión entrenados sobre subconjuntos aleatorios del conjunto de entrenamiento, empleando la técnica bootstrap y la selección aleatoria de variables en cada partición de nodo. Esta estrategia permitió minimizar el sobreajuste y mejorar la capacidad de generalización del modelo.

El desarrollo del algoritmo Random Forest se implementó con los siguientes pasos:

- Se generaron $M=400$ árboles de decisión, utilizando muestreo con reemplazo (bootstrap) sobre el conjunto de entrenamiento
- En cada nodo de cada árbol, se seleccionó aleatoriamente un subconjunto de variables candidatas para decidir la mejor división. En este caso se fijó $m=\sqrt{p}$, donde p es el número total de variables predictoras, siguiendo la heurística común para la clasificación.
- La función de ganancia de la información usada para seleccionar la división óptima fue la entropía, definida como:

$$H(S) = -\sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

Donde:

S representa el conjunto de muestras en el nodo.

c el número de clases.

p_i la proporción de la clase i en el nodo

1. Cada árbol de decisión creció hasta una profundidad máxima de 15 niveles, permitiendo capturar interacciones de segundo y tercer orden entre variables, sin generar sobreajuste extremo gracias al proceso de agregación (bagging).
2. La predicción final del bosque se obtuvo por votación mayoritaria, es decir, la clase más frecuente entre las predicciones de los 400 árboles.

$$\hat{y} = mode(\{h_m(x)\}_{m=1}^{400}) \quad (2)$$

Las variables utilizadas inicialmente fueron 20, y tras un análisis de importancia se priorizaron las siguientes:

- Antigüedad del cliente
- Tipo de persona
- Estado del cliente
- Ingresos, egresos, activos, pasivos
- Conocimiento en inversiones
- Perfil de riesgo
- Número de negociaciones realizadas
- Tipo de instrumento financiero
- Volumen negociado
- Rendimiento
- Plazo del instrumento
- Calificación de riesgo

El desempeño de este modelo fue evaluado con métricas como:

- Precision (accuracy)

$$\text{Accuracy: } \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

- TP (True Positives): Número de clientes que realmente aceptaron la oferta y fueron correctamente clasificados como tales por el modelo.
- TN (True Negatives): Número de clientes que realmente rechazaron la oferta y fueron clasificados como rechazadores por el modelo.
- FP (False Positives): Número de clientes que en realidad rechazaron la oferta, pero el modelo predijo erróneamente que aceptarían (error tipo 1).
- FN (False Negatives): Número de clientes que en realidad aceptaron la oferta, pero el modelo predijo erróneamente que la rechazan (error tipo 2).

- Recall (sensibilidad)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

- TP: Clientes realmente interesados que fueron correctamente identificados por el modelo.
 - TP + FN: Todos los clientes que en realidad aceptaron la oferta, tanto los bien clasificados como los mal clasificados.
- F1-score: Calcula la media armónica, que penaliza situaciones donde uno de los dos valores es muy bajo y equilibra la capacidad de acertar en general (precisión) y capturar la mayor parte de los positivos reales (recall).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

- Área bajo la curva ROC (AUC)

$$AUC = \int_0^1 TPR(FPR^{-1}(t))dt \quad (6)$$

La curva ROC compara:

- TPR (True Positive Rate): Tasa de verdaderos positivos

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

- FPR (False Positive Rate): Tasa de falsos positivos

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

El área bajo esa curva (AUC) mide la capacidad del modelo de distinguir entre positivos y negativos en todos los posibles umbrales de probabilidad. Mientras que, un valor de AUC cercano a 1, significa que el modelo separa perfectamente las dos clases.

2.3.2 Modelo no supervisado: K-means con MCA (Análisis de Correspondencia Múltiples)

El modelo K-means fue utilizado para segmentar la cartera de clientes en grupos homogéneos, optimizando la varianza intra-clúster y maximizando la distancia inter-clúster. El criterio matemático optimizado en K-means es la minimización de la suma de las distancias cuadráticas entre los puntos y el centroide de su clúster.

$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} ||x - u_i||^2 \quad (9)$$

- Donde:
 - S_i es el conjunto de instancias asignadas al cluster i.
 - u_i es el centroide del cluster i.
 - k es el número total de clusters a formar.

Para determinar el valor óptimo de k, se aplicó el método del codo (elbow method), que consistió en analizar la gráfica de inercia (within-cluster sum of squares) frente a distintos valores de k, identificando el punto donde la reducción de la inercia comenzaba a ser marginal.

Para determinar la calidad de los grupos generados, se valoraron métricas como:

- Puntuación de Silhouette:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (10)$$

Donde:

- a(i) es el promedio de la distancia entre el punto i y los otros puntos de su propio cluster (medida de cohesión interna).
- b(i) es la distancia promedio del punto i al cluster vecino más cercano (medida de separación externa).
- Varianza Intra-cluster (WCSS):

$$WCSS = \sum_{i=1}^k \sum_{x \in S_i} \|x - u_i\|^2 \quad (11)$$

- k es el número de clusters
- S_i es el conjunto de puntos asignados al cluster i
- u_i es el centroide del cluster i
- $\|x - u_i\|^2$ es la distancia cuadrática de cada punto al centroide de su grupo.
- Separación inter-cluster:

$$Dist(u_i, u_j) \quad (12)$$

- Para todos los pares de centroides.
- u_i, u_j : centroides de los clusters i y j.

2.5 VALIDACIÓN EXTERNA

Para validar los resultados, se realizaron entrevistas semiestructuradas con asesores financieros y jefes de cartera de la casa de valores, con el objetivo de contrastar los hallazgos del

modelo con la experiencia práctica y evaluar la coherencia de los segmentos identificados con las estrategias comerciales de la institución.

2.6 JUSTIFICACIÓN DEL ENFOQUE METODOLÓGICO

En el presente proyecto se aplicaron dos algoritmos de aprendizaje automático, correspondientes a dos enfoques metodológicos distintos: aprendizaje supervisado (Random Forest) y aprendizaje no supervisado (K-means). La elección de estos modelos responde a la naturaleza de los objetivos planteados, los cuales incluyen tanto la predicción de la aceptación de ofertas comerciales (clasificación), como la segmentación de perfiles de clientes (agrupamiento), con el propósito de proporcionar una visión integral y robusta del comportamiento de los clientes en las casas de valores.

En cuanto al aprendizaje supervisado, se utilizó el algoritmo Random Forest, debido a su alta capacidad de generalización y su resistencia al sobreajuste, característica que se deriva de la técnica de ensamble (bagging) que combina múltiples árboles de decisión. Este modelo permite identificar los factores más relevantes que explican la probabilidad de aceptación de nuevos productos financieros, lo cual resulta fundamental para la personalización de estrategias comerciales.

Por su parte, en el enfoque no supervisado se aplicó el modelo K-means, con el objetivo de segmentar la cartera de clientes en grupos homogéneos. Esta técnica, basada en la minimización de la varianza intra-cluster, destaca por su interpretabilidad y eficacia para estructurar el mercado en perfiles claramente diferenciados, lo que facilita la adaptación de ofertas comerciales según las características de cada grupo.

CAPÍTULO 3

3. Resultados

3.1 CARACTERIZACIÓN GENERAL DEL CONJUNTO DE DATOS:

3.1.1 Descripción del dataset

El conjunto de datos utilizado en este estudio comprende un total de 31.436 registros, los cuales representan transacciones bursátiles individuales realizadas por clientes —tanto personas naturales como jurídicas— en el mercado de valores ecuatoriano. Cada registro incluye información asociada a 43 variables, entre ellas: monto a invertir, rendimiento, plazo, tipo de instrumento financiero, tipo de operación, perfil de riesgo, entre otras.

3.1.2 Análisis de valores nulos y tratamiento aplicado

Durante el proceso de limpieza y análisis exploratorio, se identificó la presencia de valores nulos en cuatro variables:

- Fecha de emisión y fecha de vencimiento: Estas variables presentan 3.120 valores nulos. Esta ausencia es estructural y se explica por la naturaleza de ciertos instrumentos financieros, tales como acciones, notas de crédito, certificados de aportación, cuotas de participación, valores de contenido crediticio y valores de titularización de participación, los cuales no tienen fechas definidas de emisión o vencimiento, ya que su duración depende del tiempo que el inversionista decida mantener el instrumento.
- Rendimiento y tasa de interés: Estas variables presentan 7.125 registros faltantes. La razón es similar: algunos instrumentos no generan una rentabilidad mediante una tasa explícita, como es el caso de las acciones y ciertos valores de contenido crediticio, cuya rentabilidad se manifiesta en la apreciación del capital o en el pago de dividendos.

Se consideraron diferentes estrategias para el tratamiento de los valores faltantes. Si bien una alternativa era eliminar los registros incompletos, esto habría significado descartar una

porción relevante de instrumentos válidos en el análisis, afectando la representatividad del conjunto de datos. Por lo tanto, se optó por imputar estos valores con ceros, decisión que no introduce sesgos, dado que la ausencia no responde a errores de captura, sino a una característica intrínseca de los instrumentos financieros analizados.

3.2 RESULTADOS DEL ANÁLISIS EXPLORATORIO DE DATOS (EDA)

3.2.1 Detección y tratamiento de valores atípicos

En el análisis exploratorio se detectó valores atípicos en la mayoría de las variables cuantitativas, las cuales se presentan en la *Tabla 1*:

Tabla 1

Valores atípicos de las variables

Monto a Invertir	329
Plazo Remanente	249
Rendimiento	7
Tasa Interés	1
Valor Efectivo	270
Interes transaccional	292
Valor Anualizado	146
Total operación	252
Precio neto	10

Nota. La tabla representa los valores atípicos que presentan las variables predictoras.

Se observa que la variable con mayor cantidad de valores atípicos es Monto a invertir, con un total de 329 observaciones fuera del rango esperado, otras variables que presentan un comportamiento similar con Plazo Remanente, Valor Efectivo, Interés Transaccional y Total Operación, las cuales registran en promedio 250 valores atípicos cada una, una de las opciones

para trabajar con estos valores atípicos era reemplazarlos por el valor de cero, sin embargo, se optó por eliminar estos datos ya que al ser pocos no afectan de manera significativa al modelo, además de que con ello se disminuye la posibilidad de presentar sesgo en el modelo.

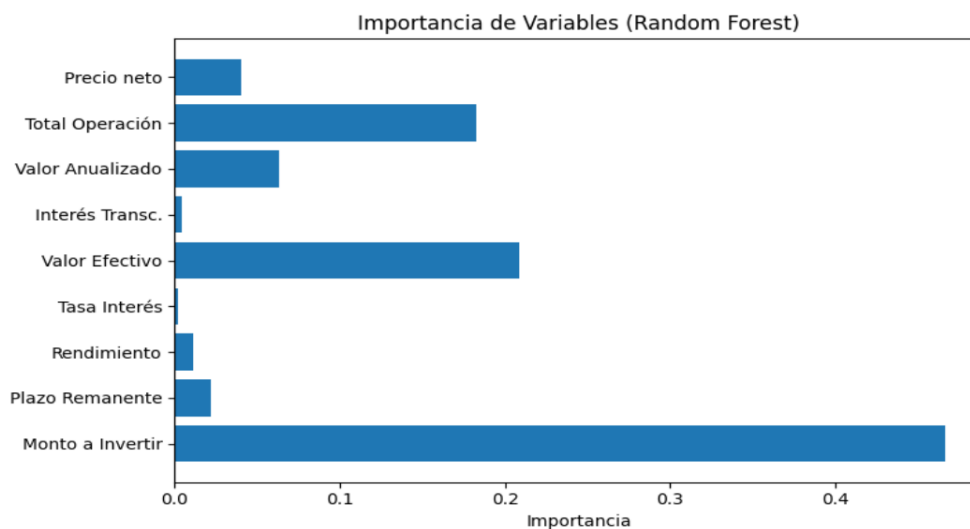
3.3 RESULTADOS DEL MODELO SUPERVISADO: RANDOM FOREST

3.3.1 Variables seleccionadas tras análisis de importancia

El modelo Random Forest fue entrenado para predecir la combinación más adecuada de tipo de producto financiero y monto de inversión, con base en los patrones históricos de comportamiento de los clientes. La *Figura 1* presenta el análisis de importancia de variables, donde se observa que el "Monto a Invertir" constituye el principal predictor del modelo, con una contribución superior al 45%. Esta variable es clave para determinar no solo el volumen recomendado, sino también el tipo de instrumento que se ajusta mejor al perfil operativo y financiero del cliente.

Figura 1

Importancia de las variables (Random Forest)



Nota. La gráfica representa la importancia de las variables predictoras (input) en un modelo Random Forest.

Asimismo, variables como el "Valor Efectivo" y el "Total de la Operación" también tienen un peso relevante, lo cual indica que los montos históricos movilizados en operaciones previas son determinantes para generar una recomendación precisa. En contraste, variables como el "Rendimiento", "Plazo Remanente" o "Interés Transcurrido" tienen una contribución marginal al modelo. Esto sugiere que, en el proceso de toma de decisiones, los aspectos técnicos del instrumento son menos relevantes que el historial transaccional y la capacidad económica del cliente.

3.3.2 Rangos de clase

Para facilitar la caracterización de los clientes en función del volumen de sus operaciones, se establecieron tres clases según el valor nominal negociado en productos financieros. Estas categorías permitieron identificar patrones diferenciados de comportamiento entre clientes con bajos, medios y altos niveles de inversión, lo cual es especialmente útil para la segmentación y análisis predictivo. La *Tabla 2* muestra los rangos establecidos para cada clase, con base en los valores mínimos y máximos observados en el conjunto de datos.

Tabla 2

Clases de inversionistas

Clase	Valor Nominal en base a la negociación de productos financieros
Bajo	(0,15 - 41912,63)
Medio	(41943,70 - 300000,00)
Alto	(300200,00 - 500000,00)

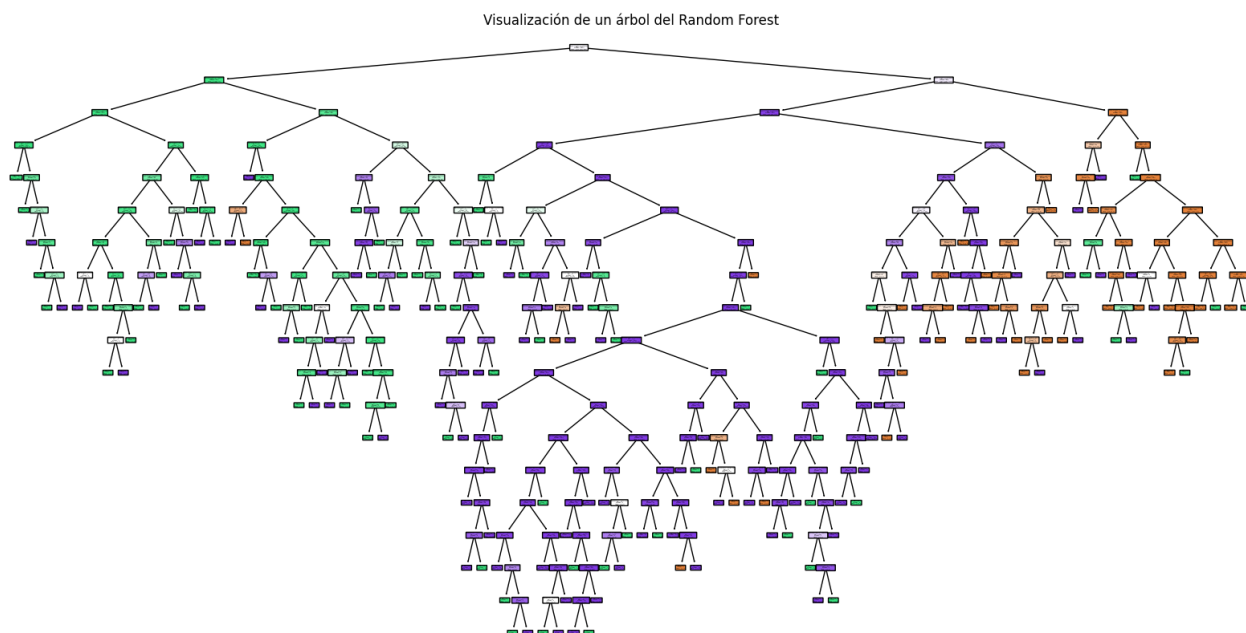
Nota. Valores mínimos y máximos por clase en el modelo.

Para llevar a cabo esta clasificación, se implementó un modelo de Random Forest, el cual resultó particularmente adecuado para el tipo de datos utilizados, por su capacidad para manejar variables heterogéneas, detectar relaciones no lineales y brindar interpretabilidad a través de los árboles de decisión que lo componen.

La visualización del árbol de decisión revela múltiples niveles jerárquicos, lo que indica que el modelo ha generado diversas reglas de decisión para clasificar las observaciones según las variables disponibles. Esta estructura sugiere una complejidad moderada, permitiendo al modelo capturar relaciones no lineales entre las variables independientes y la variable objetivo.

Figura 2

Árbol de Random Forest



Nota. La figura muestra la profundidad del árbol de decisión.

Las primeras divisiones del árbol corresponden a las variables más relevantes, según el análisis de importancia de variables, lo que refuerza su valor predictivo. Los colores de los nodos representan las distintas clases predichas, y su distribución uniforme a lo largo del árbol sugiere

que el modelo logra mantener un buen equilibrio entre las clases, evitando una sobre-representación de una categoría específica.

3.3.3 Evaluación del Desempeño del Modelo

El modelo desarrollado presenta un accuracy de 0.9951, lo que implica que aproximadamente el 99.5 % de las predicciones realizadas por el modelo son correctas. Este alto nivel de precisión general refleja una adecuada capacidad de generalización del modelo al clasificar correctamente la mayoría de los perfiles de inversión.

Además del accuracy, se incorporó el análisis de la matriz de confusión, la cual permite visualizar el número de aciertos y errores cometidos por el modelo en cada clase específica (Bajo, Medio y Alto). Este análisis detallado facilita la identificación de posibles sesgos de clasificación y el comportamiento del modelo frente a cada categoría.

Tabla 3

Matriz de confusión

Clases (Prediccion / Real)	Bajo	Medio	Alto
Bajo	1778	1	2
Medio	1	1791	7
Alto	3	13	1953

Nota. La tabla muestra el desempeño del modelo en términos de aciertos y errores de clasificación.

En el caso de la clase Medio, el modelo acertó en 1791 predicciones, aunque clasificó erróneamente 1 observación de la clase Bajo y 13 observaciones de la clase Alto como pertenecientes a esta clase. Finalmente, para la clase Alto, se obtuvieron 1953 predicciones correctas, mientras que se incurrió en 2 errores clasificando observaciones de la clase Bajo y 7 de la clase Medio como si fueran de la clase Alto. Estos resultados reflejan un alto nivel de precisión,

ya que la mayoría de los valores se concentran en la diagonal de la matriz, lo cual indica que el modelo logra identificar adecuadamente el perfil de inversión en la mayoría de los casos.

A continuación, se presenta el reporte de clasificación, que detalla métricas específicas por clase:

Figura 3

Reporte de clasificación

Reporte de clasificación				
	Precisión	Recall	F1-score	Support
Alto	1,00	1,00	1,00	1781
Bajo	0,99	1,00	0,99	1799
Medio	1,00	0,99	0,99	1969
Accuracy			1,00	5549
Macro avg	1,00	1,00	1,00	5549
Weighted avg	1,00	1,00	1,00	5549

Nota: La tabla presenta las métricas de evaluación del modelo para cada clase.

- **Clase de Inversión: Alto**

Para la clase Alto, el modelo mostró un rendimiento impecable en todas las métricas (precisión, recall y F1-score), alcanzando el 100 % en cada una, lo que indica que todas las observaciones fueron correctamente clasificadas sin errores.

- **Clase de Inversión: Medio**

Para la clase Medio, el modelo presentó un desempeño muy alto, con una precisión del 100 %, lo que indica que no hubo falsos positivos, y un recall del 99 %, reflejando una ligera pérdida en la identificación completa de los casos reales. El F1-score de 99 % confirma un rendimiento consistente y equilibrado, con un margen de error mínimo.

- **Clase de Inversión: Bajo**

La clase Bajo fue identificada con precisión del 99 % y recall del 100 %, reflejando un desempeño sólido y balanceado, con un mínimo margen de error en predicciones positivas.

Estas métricas reflejan que el modelo Random Forest tiene un desempeño robusto y confiable en la predicción del perfil de inversión de los clientes, lo cual es clave para una correcta segmentación y toma de decisiones financieras estratégicas en la casa de valores.

3.3.4 Interpretación de resultados: factores que predicen el volumen a negociar del cliente

Este enfoque de clasificación no solo permite segmentar a los clientes según su perfil de negociación, sino también comprender cuáles son los factores determinantes en su comportamiento financiero, aspecto crucial para casas de valores cuyo principal grupo de clientes está compuesto por personas jurídicas. El modelo Random Forest reveló que variables como el Valor Efectivo y el Total de la Operación tienen un peso significativo en la predicción del volumen negociado, lo que indica que los montos históricos movilizados en operaciones previas son los principales impulsores en la asignación de clases. En cambio, variables técnicas relacionadas con el instrumento financiero, como el Rendimiento, el Plazo Remanente o el Interés Transcurrido, tuvieron una influencia marginal.

Tabla 4*Importancia de variables*

Importancia de Variables	
VARIABLES	PORCENTAJE
Precio neto	0,04
Total Operación	0,18
Valor Anualizado	0,07
Valor Efectivo	0,22
Monto a Invertir	0,46

Nota: La tabla presenta el peso de las variables principales en el modelo.

Este resultado sugiere que, en el contexto del mercado bursátil ecuatoriano, la capacidad económica del cliente y su historial transaccional son factores mucho más relevantes que las características técnicas del producto al momento de tomar decisiones de inversión. Esta evidencia resulta especialmente útil para que las casas de valores ajusten su estrategia comercial, pudiendo segmentar a los asesores según el tipo de cliente y su clase asignada, y ofrecer productos personalizados y más relevantes según el perfil identificado.

En este sentido, el uso de machine learning no solo mejora la eficiencia y precisión del análisis, sino que se convierte en una herramienta estratégica para optimizar la colocación de ofertas, fortalecer la fidelización y mejorar la asignación de recursos comerciales, permitiendo a la organización responder con mayor agilidad y enfoque a las necesidades del mercado.

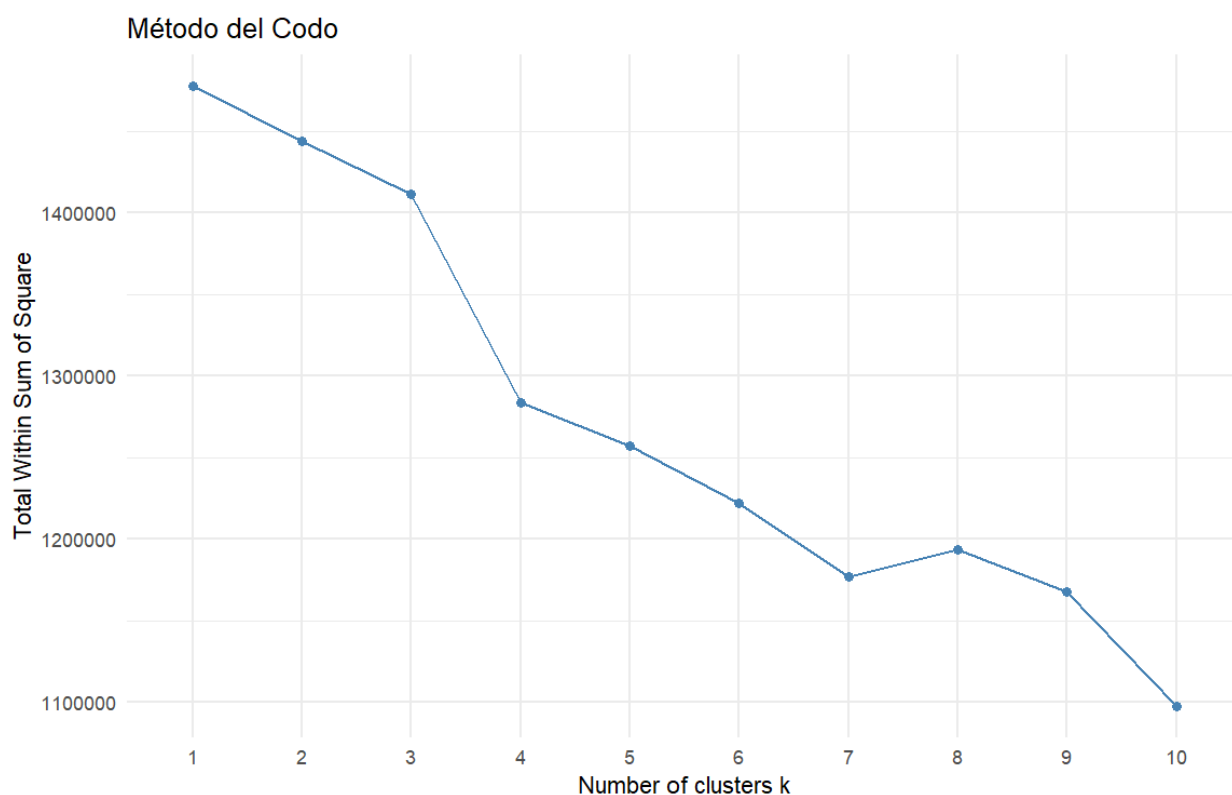
3.4 RESULTADOS DEL MODELO NO SUPERVISADO: K-MEANS

3.4.1 Selección del número óptimo de clústers

Para determinar el número óptimo de agrupamientos (clúster), se aplicó el método del codo, que consiste en graficar la suma total de los errores cuadráticos dentro del grupo (Within-Cluster Sum of Squares, WCSS) frente al número de clúster (K). El punto en el que la disminución del WCSS comienza a volverse menos pronunciada indica el valor óptimo de K, ya que representa un equilibrio entre simplicidad del modelo y capacidad explicativa.

Figura 4

Método del Codo



Nota. Se realizó este gráfico con el fin de obtener el número de clústers óptimos.

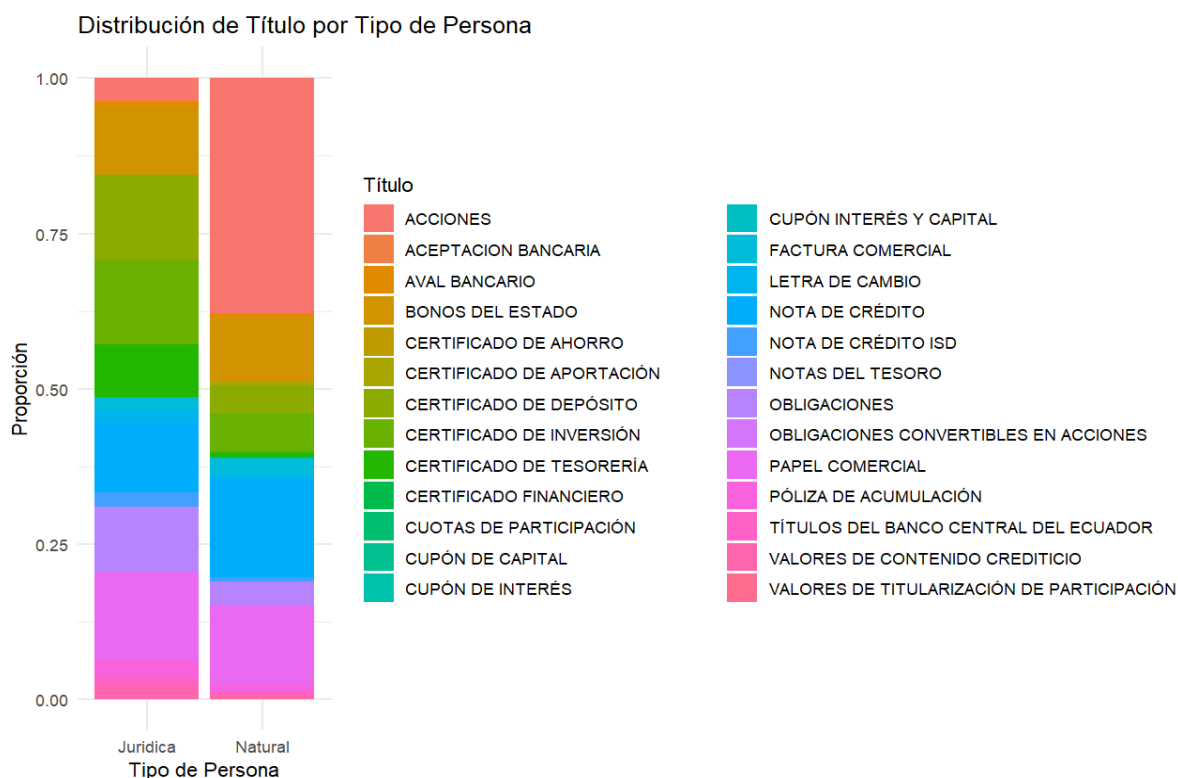
Como se observa en la *Figura 4*, el punto de inflexión más evidente ocurre entre $K = 2$ y $K = 4$, siendo el valor $K = 2$ el seleccionado para este análisis. Esta elección se fundamenta en que, a partir de este punto, la reducción en la varianza intra-clúster comienza a estabilizarse, lo que sugiere que dos agrupaciones representan de forma eficiente la estructura subyacente de los datos sin introducir complejidad innecesaria al modelo.

3.4.2 Caracterización de los clústers: perfil del cliente por grupo

Tras aplicar el algoritmo K-means con $K = 2$, se identificaron dos segmentos bien definidos de clientes, diferenciados principalmente por el tipo de persona, el tipo de instrumento que negocian y el nivel de diversificación de su portafolio.

Figura 5

Distribución del Título por Tipo de Persona

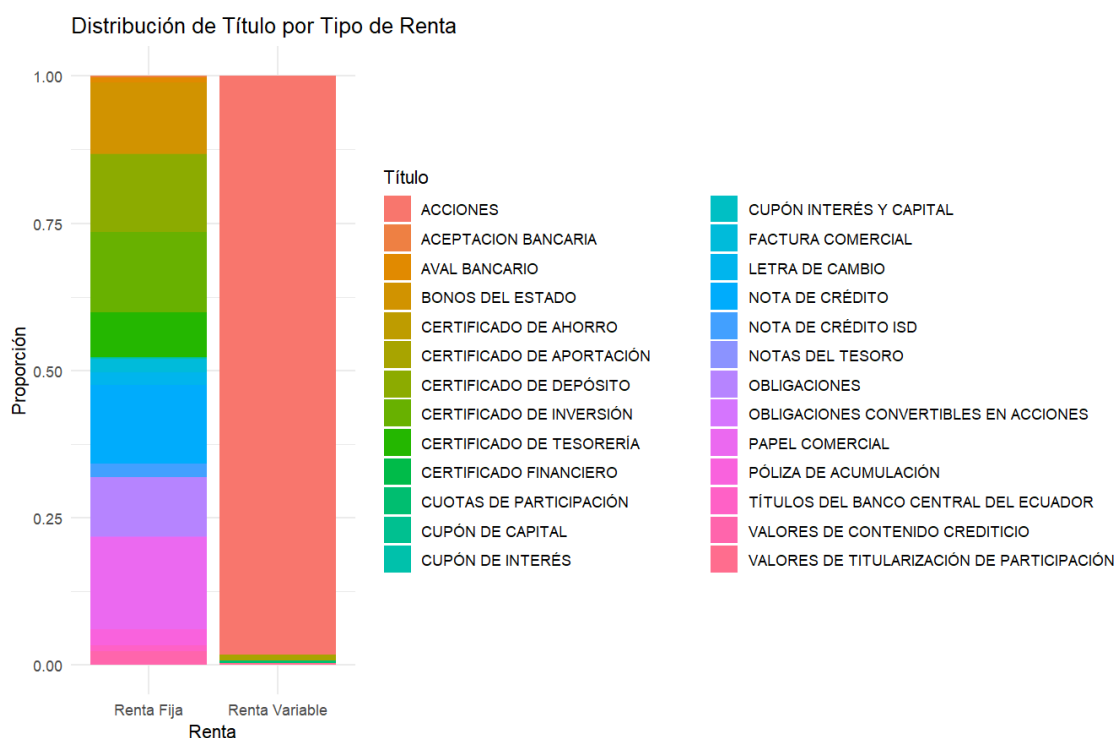


Nota. Se observa que las acciones son el instrumento financiero mayormente obtenido por los tipos de clientes naturales

La *Figura 5* muestra la distribución de los títulos financieros según el tipo de persona (jurídica o natural). Se observa que los clientes jurídicos presentan una mayor diversidad en los instrumentos negociados, con alta participación en obligaciones, pólizas de acumulación, certificados financieros y valores de contenido crediticio. Esto sugiere un mayor grado de sofisticación y estrategias más estructuradas. Por otro lado, los clientes naturales concentran su participación en productos más tradicionales como acciones, bonos del Estado y letras de cambio, lo cual refleja decisiones más simples o conservadoras.

Figura 6

Distribución del Título por Tipo de Renta



Nota. Se observa que las acciones son las que mayor porcentaje representan del tipo de productos financieros en renta variable

Adicionalmente, la *Figura 6* analiza la distribución de títulos según el tipo de renta (fija o variable). La mayoría de los instrumentos negociados corresponden a renta fija, lo cual indica

una preferencia general del mercado por instrumentos estables y de menor riesgo. Sin embargo, en el segmento de renta variable, las acciones son el producto predominante, y este tipo de instrumentos tiende a ser más común entre los clientes naturales, lo que refuerza la distinción de perfiles observada en el clúster.

A partir de este análisis, se caracterizan los grupos obtenidos de la siguiente manera:

Tabla 5

Clustering

Clúster	Tipo de Cliente	Tipo de Renta	Diversificación	Instrumentos comunes
0	Jurídico	Fija y Variable	Alta	Obligaciones, pólizas, certificados, etc.
1	Natural	Variable y Fija	Media-Baja	Acciones, bonos del estado, etc.

Nota. La tabla muestra las características de los grupos obtenidos del análisis anterior

Esta caracterización resulta fundamental para comprender el comportamiento de los distintos perfiles de cliente. El clúster 1 agrupa a clientes institucionales o empresariales con un enfoque estructurado, mientras que el clúster 2 refleja perfiles individuales con estrategias más simples y con distintas exposiciones al riesgo. Estos hallazgos permitirán que las casas de valores diseñen estrategias diferenciadas, asignen asesores especializados según el perfil y ajusten la promoción de productos de forma más eficiente y personalizada.

3.4.3 Interpretación de los grupos: diferencias clave entre perfiles

El análisis no supervisado mediante K-means permitió identificar dos perfiles bien diferenciados de clientes en el mercado de valores, cuya segmentación aporta información valiosa para la estrategia comercial de las casas de valores con el fin de ofrecer productos financieros personalizados a sus clientes.

El primer grupo (clúster 1) está conformado principalmente por personas jurídicas, con un portafolio altamente diversificado y una preferencia clara por instrumentos de renta fija, tales como obligaciones, certificados financieros, pólizas de acumulación y valores estructurados. Estos clientes suelen realizar operaciones más sofisticadas, con montos más altos y objetivos financieros más definidos, lo que sugiere una estrategia orientada a la preservación de capital y estabilidad de retorno. Su comportamiento refleja un perfil institucional que requiere asesoría especializada, productos estructurados y soluciones adaptadas a las necesidades corporativas.

Por otro lado, el segundo grupo (clúster 2) agrupa en su mayoría a personas naturales, con una menor diversificación en sus inversiones y una combinación a instrumentos de renta variable y fija, especialmente acciones, obligaciones y papel comercial. Este segmento muestra un comportamiento más conservador o especulativo, posiblemente impulsado por decisiones individuales, menor nivel de asesoramiento financiero o acceso limitado a productos complejos.

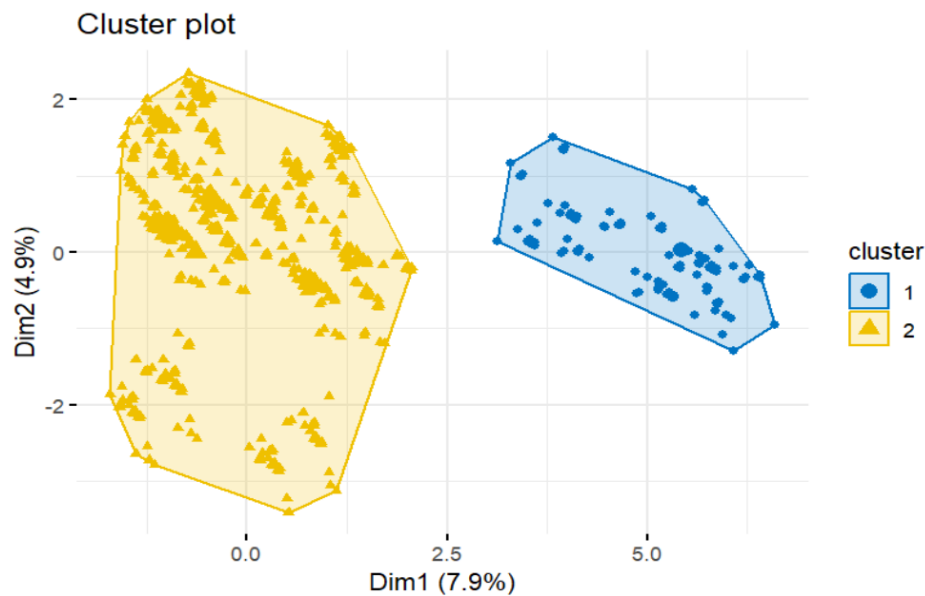
La distinción entre estos grupos permite a las casas de valores diseñar estrategias diferenciadas según el tipo de cliente:

- Para clientes del clúster 1, se puede priorizar la asignación de asesores especializados, desarrollo de productos de inversión corporativos y gestión activa de relaciones comerciales.
- Para clientes del clúster 2, resulta clave implementar estrategias de educación financiera, recomendaciones automatizadas y productos estandarizados de menor riesgo.

Modelo K-means con MCA (Análisis de correspondencia Múltiples)

Figura 7

Clúster Plot



Nota: La figura muestra la separación de grupos en el modelo

La *Figura 3* representa gráficamente los resultados del algoritmo de clustering K-means con $K = 2$, proyectados sobre dos dimensiones principales obtenidas mediante reducción de dimensionalidad. En este gráfico cada punto representa un cliente, y los colores distinguen los dos grupos generados:

- **clúster 1:** Agrupa a clientes que presentan características similares entre sí, pero diferenciadas claramente del otro grupo, se puede observar que este grupo tiende a tener una dispersión más controlada, con lo cual podemos decir que presenta homogeneidad interna en sus variables clave.

- **clúster 2:** Representa un grupo más numeroso y disperso, lo cual indica una mayor diversidad dentro del grupo, el gráfico muestra una clara separación entre el mismo y el clúster 1 con una frontera clara entre ambos segmentos.

La separación visual entre ambos clústers indica que el modelo ha logrado una segmentación efectiva, agrupando clientes con perfiles distintos. Esta separación valida que las variables utilizadas para el clustering, tales como: tipo de persona, categoría, título, sector, renta y tipo de operación, permiten distinguir dos comportamientos de inversión bien definidos en el mercado bursátil ecuatoriano.

Esta visualización del modelo K-means respalda la interpretación anterior:

- El clúster 1 corresponde principalmente a clientes jurídicos, con portafolios más estructurados y preferencias hacia productos de renta fija.
- El clúster 2 agrupa a clientes naturales, con estrategias más simples y una combinación de uso de instrumentos de renta variable y de renta fija.

Tabla 6

Distribución de productos financieros en cada grupo

Clúster	Título	n	Porcentaje
2	ACCIONES	2834	26,27
2	CERTIFICADO DE APORTACIÓN	32	0,3
2	CUOTAS DE PARTICIPACIÓN	12	0,11
2	CUPÓN DE CAPITAL	20	0,19
2	CUPÓN DE INTERÉS	19	0,18
2	CUPÓN INTERÉS Y CAPITAL	4	0,04
2	FACTURA COMERCIAL	639	5,92

2	OBLIGACIONES	2822	26,16
2	PAPEL COMERCIAL	4287	39,74
2	VALORES DE CONTENIDO CREDITICIO	106	0,98
2	VALORES DE TITULARIZACIÓN DE PARTICIPACIÓN	13	0,12
1	ACCIONES	423	2,05
1	ACEPTACIÓN BANCARIA	68	0,33
1	AVAL BANCARIO	232	1,12
1	BONOS DEL ESTADO	3390	16,42
1	CERTIFICADO DE AHORRO	38	0,18
1	CERTIFICADO DE DEPÓSITO	3734	18,08
1	CERTIFICADO DE INVERSIÓN	3825	18,52
1	CERTIFICADO DE TESORERÍA	2139	10,36
1	CERTIFICADO FINANCIERO	4	0,02
1	CUPÓN DE CAPITAL	24	0,12
1	LETRA DE CAMBIO	605	2,93
1	NOTA DE CRÉDITO	3770	18,26
1	NOTA DE CRÉDITO ISD	626	3,03
1	NOTAS DEL TESORO	3	0,01
1	OBLIGACIONES	28	0,14
1	OBLIGACIONES CONVERTIBLES EN ACCIONES	46	0,22
1	PAPEL COMERCIAL	83	0,4

1	PÓLIZA DE ACUMULACIÓN	759	3,68
1	TÍTULOS DEL BANCO CENTRAL DEL ECUADOR	314	1,52
1	VALORES DE CONTENIDO CREDITICIO	537	2,6

Nota: La tabla muestra la distribución de títulos de los productos financieros y los valores para tanto para cada clúster como para cada título

La *Tabla 6* muestra la distribución de productos financieros negociados por los clientes en cada uno de los dos clústers identificados mediante el algoritmo K-means. Este análisis permite caracterizar los intereses y comportamientos financieros de cada grupo, en función de los títulos que prefieren negociar.

Clúster 1: Clientes de tipo jurídico

El clúster 1, que representa aproximadamente el 70 % de la muestra, se caracteriza por una alta concentración en instrumentos de tipo de renta fija como:

- Bonos del estado: 16,41 %
- Certificados de depósito: 18,08 %
- Certificado de inversión: 18,51%
- Notas de crédito: 18,16%

Estos productos están asociados a clientes jurídicos, lo cual coincide con el análisis previo realizado basado en el tipo de persona. Este grupo también participa en títulos como facturas comerciales, cuotas de participación y demás instrumentos, lo que indica un enfoque de inversión más diversificado, pero con productos de menor complejidad.

Clúster 2: Clientes de tipo natural

El clúster 2 tiene un comportamiento diferente al clúster 2, ya que presenta mayor participación en instrumentos financieros típicamente utilizados por personas naturales, tales como:

- Acciones: 26,27 %
- Obligaciones: 26,26 %
- Papel comercial: 39,64 %

Esta distribución en el clúster refleja una preferencia por instrumentos de renta fija con un aproximado del 65%, asociados con una estrategia más conservadora, y con un aproximado del 26% de adquisición en instrumentos financieros de renta variables, con una estrategia financiera arriesgada con cambios constantes en el mercado. También se observa participación en instrumentos estructurados como valores de contenido crediticio, pólizas de acumulación, papel comercial y facturas comerciales, lo que sugiere un mayor grado de sofisticación y análisis financiero.

3.4.4 Métricas de Evaluación

- **Puntuación media de Silhouette:** 0.51

Tabla 7

Distribución de productos financieros en cada grupo

clúster	size	ave.sil.width
1	10788	0.55
2	10648	0.47

Nota: La tabla muestra el tamaño de las observaciones en cada clúster y la puntuación media de Silhouette

La puntuación media de Silhouette obtenida del modelo fue de 0.51, lo que indica una estructura de clustering razonablemente buena. Esta métrica evalúa qué tan bien están separados los grupos formados por el algoritmo de clustering y qué tan coherentes son internamente.

Por otra parte, la puntuación media de Silhouette de 0.51 sugiere que la mayoría de los clientes están adecuadamente agrupados, y que los dos clústers generados presentan una separación aceptable entre sí, lo cual se puede verificar en la *Figura 7* realizado para los clústers anteriormente presentados. Esta puntuación respalda que la elección de $K = 2$ fue adecuada y que el modelo logra capturar estructuras reales en los datos, lo que valida el análisis de segmentación realizado.

El modelo no supervisado K-means fue evaluado utilizando dos métricas adicionales fundamentales para medir su rendimiento: la suma de varianza intra-clúster (WCSS) y la proporción de varianza explicada (BSS/TSS). Estas medidas permiten validar la coherencia interna de los grupos y la calidad de la segmentación lograda.

- **Suma de varianza intra-cluster (WCSS): 1408**

La WCSS (Within-Cluster Sum of Squares) representa la suma de las distancias cuadradas entre cada punto y el centroide de su clúster, como se explicó anteriormente en el capítulo 1, es una medida de compactación interna: cuanto más bajo es el valor de WCSS, más homogéneo es cada clúster internamente.

En este caso, una WCSS de 1408 indica que los dos clústers generados presentan una compactación razonablemente buena, es decir, los clientes asignados a cada grupo comparten características similares en términos de las variables consideradas (tipo de cliente, productos financieros, montos invertidos, tipo de operación, categoría, etc.). Este valor, combinado con una

puntuación media de Silhouette de 0.51, respalda que los grupos no solo son consistentes internamente, sino también que están bien diferenciados entre sí.

- **Separación inter-cluster (BSS/TSS): 0.547**

La métrica BSS/TSS (Between Sum of Squares / Total Sum of Squares) indica la proporción de la variabilidad total explicada por la estructura de clustering. Es decir, esta métrica mide qué tan bien el modelo separa los grupos entre sí, su valor varía entre 0 (mala separación) y 1 (separación perfecta).

El modelo K-means presenta un valor de 0.547, lo cual implica que el 54.7 % de la variación total en los datos es explicada por los dos clústers generados, lo cual es una partición sólida para un modelo no supervisado aplicado. Este resultado sugiere que el modelo logra identificar estructuras significativas en los datos y que existe una separación clara entre los perfiles de cliente captados por cada clúster.

CAPÍTULO 4

4. Conclusiones y recomendaciones

4.1 CONCLUSIONES

La combinación de ambos enfoques metodológicos, aprendizaje supervisado y no supervisado, proporciona una visión integral del comportamiento de los clientes, la cual permite optimizar la segmentación, personalización de las ofertas y el fortalecimiento de la fidelización de los clientes en la casa de valores.

La incorporación de técnicas de Machine Learning en el análisis del mercado de valores no solo incrementa la precisión y velocidad del procesamiento de datos, sino que constituye una herramienta estratégica para anticipar tendencias, reducir riesgos y aprovechar nuevas oportunidades de negocio en el sector bursátil.

- El diseño del modelo no supervisado K-means permitió identificar dos clústers diferenciados: Clientes Jurídicos, con portafolios diversificados y preferencia hacia instrumentos de renta fija como lo son los bonos del estado, certificados de depósito, certificados de inversión y notas de crédito, productos financieros que predominan en la conformación de sus carteras de inversión, a su vez, los Clientes Naturales, con menor diversificación y mayor inclinación hacia instrumentos de renta variable, especialmente acciones. La puntuación media de Silhouette (0,51) y la separación inter-cluster ($BSS/TSS = 0,547$) confirman que la segmentación lograda es coherente y efectiva. En conjunto, estos indicadores apuntan a que la segmentación obtenida no solo es estadísticamente significativa, sino que tiene potencial para ser utilizada en la toma de decisiones estratégicas, como el diseño de campañas diferenciadas, la personalización de ofertas o la priorización de recursos hacia segmentos más rentables o estratégicos.
- El análisis de importancia de variables identificó que el Monto a Invertir, el Valor Efectivo y el Total de la Operación constituyen los factores más relevantes para la predicción del

comportamiento de inversión de los productos financieros otorgados por la casa de valores como bonos del estado, acciones, bonos del estado, certificados de inversión, obligaciones y notas de crédito superando en relevancia a variables técnicas como el rendimiento, el plazo remanente o el interés transcurrido. Esto confirma que la capacidad económica y el historial transaccional del cliente son elementos determinantes para anticipar sus decisiones de inversión.

- Se ha determinado que el modelo supervisado Random Forest presenta un rendimiento sobresaliente en la predicción del perfil de inversión de los clientes del mercado de valores, alcanzando un nivel de precisión del 99,5 %. Este resultado evidencia su eficacia para clasificar a los clientes según su volumen de inversión y el tipo de producto financiero, ofreciendo un soporte sólido para la toma de decisiones estratégicas.

4.2 RECOMENDACIONES

El presente proyecto se centró en el análisis del comportamiento de los clientes que negociaban en el mercado de valores ecuatoriano, por medio del desarrollo de algoritmos de aprendizaje automático y modelos estadísticos. Este enfoque permitió identificar una serie de patrones de inversión y clasificar a los clientes en dos perfiles, lo que constituye una base sólida para diseñar estrategias diferenciadas. A partir de estos hallazgos, se presentan las siguientes recomendaciones:

- Se considera necesario que las casas de valores incorporen modelos predictivos basados en técnicas de machine learning, con el fin de proveer a los asesores comerciales de información respaldada estadísticamente. De esta manera, se fortalece la capacidad de ofrecer una asesoría más personalizada y alineada a los perfiles de inversión identificados, en el que potencia tanto la eficiencia en la gestión como la satisfacción de los clientes.

- Se sugiere implementar estrategias comerciales diferenciadas orientadas a clientes institucionales, priorizando la asignación de asesores especializados con experiencia en operaciones de alto volumen y complejidad. Asimismo, resulta clave fomentar el desarrollo de productos estructurados y soluciones de inversión adaptadas a las necesidades corporativas, junto con una gestión activa de las relaciones comerciales. Estas acciones permitirán fortalecer la confianza, optimizar la preservación del capital y asegurar la estabilidad de los retornos en este segmento.
- Dado que el segmento de clientes naturales se caracteriza por realizar un número reducido de transacciones al año y, en muchos casos, carece de un nivel profundo de asesoramiento financiero, se recomienda implementar estrategias orientadas a la educación y acompañamiento del cliente. Entre las acciones sugeridas se encuentran: la automatización de procesos informativos mediante videos tutoriales o plataformas digitales interactivas que expliquen de manera sencilla el funcionamiento del mercado de valores, así como la oferta de talleres o cursos de formación con un costo accesible. Estas iniciativas permitirán mejorar la comprensión del proceso de inversión, incrementar la confianza del cliente y fomentar una participación más activa y consciente en el mercado.
- Se recomienda actualizar periódicamente los modelos con datos recientes, asegurando su adaptación a cambios en las condiciones del mercado y en el comportamiento de los clientes. Además, es recomendable verificar de forma continua que los modelos no incurran en sesgos que afecten de manera injusta a determinados grupos, garantizando el cumplimiento de principios éticos y regulatorios en el uso de inteligencia artificial aplicada al sector financiero.
- Se propone emplear técnicas de obtención de datos externos, como el web scraping, para complementar la información interna y evaluar de manera más integral el perfil de los

clientes dentro del sistema financiero ecuatoriano. La incorporación de fuentes externas, tales como información tributaria, indicadores macroeconómicos o datos regulatorios, permitiría enriquecer los modelos predictivos, mejorando la precisión en los procesos de segmentación.

De este modo, las casas de valores no solo optimizarán la personalización de sus estrategias comerciales, sino que también fortalecerán sus procesos de cumplimiento normativo, reduciendo el riesgo de captar fondos provenientes de fuentes no permitidas por la regulación vigente.

Referencias

- Ochoa, J. (2024). Optimización de portafolios de inversión en la Bolsa de Valores de Quito, Ecuador utilizando redes neuronales artificiales (Tesis de maestría). Universidad de las Américas, Quito. <http://dspace.udla.edu.ec/handle/33000/15983>
- Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL). (2020). *Proyecto diagnóstico de inteligencia artificial en Ecuador: sector financiero*. Quito, Ecuador.
- Ochoa Luzuriaga, J. J. (2024). *Optimización de portafolios de inversión en la Bolsa de Quito mediante redes neuronales artificiales*. Universidad Internacional del Ecuador.
- Proyecto “El Mercado de Valores del Ecuador: Oportunidades y Desafíos en la Era Fintech 3.0”. (2024). *X-pedientes Económicos*.
- Alonso Robisco, A., & Carbó, J. M. (2022, 10 19). Inteligencia artificial y finanzas: una alianza estratégica. Banco de España Eurosistema, 2222(electrónica), 19.
- Mercado de valores ecuatoriano. (s. f.). <https://www.supercias.gob.ec/portalscv/Institucion/Mercado-de-Valores.php?seccion=Historia-del-Mercado-de-Valores-Ecuadoriano#:~:text=La%20primera%20Ley%20de%20Mercado,de%20compa%C3%B1%C3%ADas%20an%C3%B3nimas%20a%20corporaciones>
- Bolsa de Valores Guayaquil | Casas de Valores | ¿Qué es una Casa de Valores? (s. f.). <https://www.bolsadevaloresguayaquil.com/casas-de-valores/casas-de-valores.asp#:~:text=Las%20Casas%20de%20Valores%20son,de%20Compa%C3%B1%C3%ADas%20C%20Valores%20y%20Seguros>.
- Superintendencia de Compañías, Valores y Seguros. (1993). *Ley de Mercado de Valores*. Registro Oficial Suplemento No. 252. Quito, Ecuador. <https://www.bolsadequito.com/index.php/historia-institucional2>

- Coba, G. (2024, 26 febrero). Gobierno y empresas buscan financiamiento en un mercado de valores estancado. Primicias. <https://www.primicias.ec/noticias/economia/mercado-valores-ecuador-gobierno-bonos-estado-empresas/>
- Banco Central del Ecuador. (s.f.). *Depósito Centralizado de Valores del BCE*. <https://www.bce.fin.ec>
- Bolsa de Valores de Quito. (s.f.). *Participantes del mercado de valores*. <https://www.bvq.fin.ec>
- Decevale. (s.f.). *Quiénes somos*. <https://www.decevale.com.ec>
- Superintendencia de Compañías, Valores y Seguros. (s.f.). *Catastro Público del Mercado de Valores*. <https://www.supercias.gob.ec>
- Dútan Pugo, W. S. (2021). *Análisis del comportamiento de los mercados de valores en Ecuador (2011–2021)*. Universidad Politécnica Salesiana.
- AWS AMAZON. (n.d.). Aprendizaje supervisado frente a aprendizaje no supervisado: diferencia entre los algoritmos de machine learning. AWS. Retrieved June 12, 2025, from <https://aws.amazon.com/es/compare/the-difference-between-machine-learning-supervised-and-unsupervised/>
- Francés Monedero, T. (2020, Abril). Impacto del machine learning en el sistema financiero. Repositorio Universidad Pontificia Comillas. Retrieved 06 09, 2025, from <https://repositorio.comillas.edu/xmlui/handle/11531/42692>
- Hinestroza Ramirez, D. (2018, 06 11). El Machine Learning a través de los tiempos, y los aportes a la humanidad. CORE. Retrieved 06 09, 2025, from https://core.ac.uk/outputs/233044684/?utm_source=pdf&utm_medium=banner&utm_campaign=pdf-decoration-v1
- Laurisch, T., Stanwick, K., & Antwi, A. (2024, June 11). The ultimate guide to customer behavior analysis. Mastercard Data & Services. Retrieved June 11, 2025, from

<https://www.mastercardservices.com/en/industries/retail/insights/ultimate-guide-customer-behavior-analysis>

- SRI. (2018). PLAN DE CONTROL Y LUCHA CONTRA EL FRAUDE FISCAL. Servicio de Rentas Internas. Retrieved 06 09, 2025, from https://www.sri.gob.ec/web/intersri/buscador?p_p_id=com_liferay_portal_search_web_search_results_portlet_SearchResultsPortlet_INSTANCE_uzxPwvfxyoLa&p_p_lifecycle=0&p_p__state=maximized&p_p_mode=view&_com_liferay_portal_search_web_search_results_portlet_Sear
- Sufian Mozumder, M. A., Mahmud, F., Shujan Shak, M., Sultana, N., Rodrigues, G. N., Rafi, M. A., Rahman Farazi, M. Z., Razaul Karim, M., Khan, M. S., & Shahriar Mahmud Bhuiyan, M. (2024, Agosto). Optimizing Customer Segmentation in the Banking Sector: A Comparative Analysis of Machine Learning Algorithms. *Journal of Computer Science and Technology Studies*, 6(4), 7. 10.32996/jcsts.2024.6.4.1
- Lokanan ME y Ramzan S (2024) Predicción de dificultades financieras en empresas que cotizan en TSX utilizando algoritmos de aprendizaje automático. *Frente. Artif. Intell.* 7:1466321. doi: 10.3389/frai.2024.1466321
- González, G. A. A., Vásquez, M. G., Vélez, J. P., & Hernández, A. (2021). Big data y las implicaciones en la cuarta revolución industrial - retos, oportunidades y tendencias futuras. *Revista Venezolana De Gerencia*, 26(93), 33-47. <https://doi.org/10.52080/rvg93.04>
- Figueroa, M. A., Turner-Szymkiewicz, D., Lopez-Rojas, E. A., Cárdenas-Rodriguez, J. S., & Norinder, U. (2021). Un enfoque para hacer benchmark a los algoritmos para la detección de fraude en la era covid-19. *Revista Latinoamericana De Economía Y Sociedad Digital*, (2). <https://doi.org/10.53857/eorg4750>

- Guerrero, M. d. R. D. (2023). La seguridad de invertir en plataformas de inversión digitales en Ecuador. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, 4(1).
<https://doi.org/10.56712/latam.v4i1.556>
- Fuentes, N. L. F. R., Pupo, N. A. S., Torres, N. A. F. E., Icaza, N. V. M. S., & Angulo, N. L. E. V. (2019b). Estudio comparado de las técnicas de Inteligencia Artificial para el diagnóstico de enfermedades. *Dilemas Contemporáneos Educación Política y Valores*.
<https://doi.org/10.46377/dilemas.v30i1.1264>
- Castro, T. A. M., Fernanda, M. F. M., Gómez, J. S. A., & Barbosa, W. G. J. (2020b). Aplicaciones móviles de salud con respaldo en publicaciones científicas, para la mejora en el ejercicio de la optometría y la oftalmología. *Ciencia y Tecnología Para la Salud Visual y Ocular*, 17(2), 51-63. <https://doi.org/10.19052/sv.vol17.iss2.6>
- Castaneda, A. U. (2023b). Un viaje hacia la inteligencia artificial en la educación. *Realidad y Reflexión*, 56, 121-136. <https://doi.org/10.5377/ryr.v1i56.15776>
- Verduga, J. E. B., Casanova, J. C. A., & Lara, A. R. (2024b). Inteligencia Artificial aplicada al área comercial de la empresa constructora-inmobiliaria PARMAROT S.A. *Polo del Conocimiento*, 9(4), 2849-2869. <https://doi.org/10.23857/pc.v9i4.7094>
- Denniye, H. R. (2018b). El Machine Learning a través de los tiempos, y los aportes a la humanidad. *CORE*.
https://core.ac.uk/outputs/233044684/?utm_source=pdf&utm_medium=banner&utm_campaign=pdf-decoration-v1