

**Pronóstico del comportamiento de los activos administrados por los fondos de inversión
colectiva de las sociedades fiduciarias en Colombia mediante modelos de inteligencia
artificial y herramientas econométricas**

Omar Andrés Montañez Murillo

Asesor

Andrés Felipe Solis Pino

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Dedicatoria

A Dios, por ser mi guía, mi fortaleza y fuente de inspiración en cada etapa de este camino.
Gracias por darme la paciencia y la fuerza para superar este desafío. A mi esposa, por su amor incondicional, su apoyo constante y su comprensión. Gracias por ser mi refugio y motivación en cada momento.

Agradecimientos

Expreso mi más sincero agradecimiento a todas las personas e instituciones que, de una u otra forma, contribuyeron a la realización de este trabajo de grado.

En primer lugar, agradezco a Andrés Felipe Solís Pino, por su invaluable guía, orientación y apoyo a lo largo de todo el proceso investigativo. Su conocimiento y experiencia fueron fundamentales para la consolidación de este estudio.

A Darwin Martínez, quien con sus consejos y colaboración hicieron de este proceso una experiencia enriquecedora. Su compañía y apoyo fueron esenciales para el desarrollo de esta investigación.

A mi familia, por su constante apoyo, paciencia y motivación incondicional. Gracias por ser mi fuente de inspiración y por acompañarme en cada paso de este camino, brindándome el ánimo necesario para superar los desafíos.

A la Asociación de Fiduciarias, por proporcionar los recursos, datos y conocimientos necesarios para la realización de este trabajo.

Finalmente, a todos aquellos que, de manera directa o indirecta, contribuyeron con su tiempo, conocimientos y apoyo en la culminación de esta etapa académica. A cada uno de ustedes, mi más sincero reconocimiento y gratitud.

Resumen

Los Fondos de Inversión Colectiva se han consolidado como instrumentos de inversión destacados, administrando un promedio de \$72 billones en los últimos cinco años. Estos fondos han demostrado una tasa de crecimiento anual compuesta del 6.35% y un incremento anual promedio del 10% en el número de inversionistas. A pesar de estos indicadores positivos y de ser productos insignia en términos de rendimientos, en Colombia no se han desarrollado estudios de modelado que proporcionen herramientas prospectivas para prever el comportamiento de los activos administrados. Tal previsión podría ofrecer ventajas significativas, como anticiparse a periodos de escasez de liquidez.

En respuesta a esta necesidad, el objetivo de este proyecto es desarrollar las fases necesarias para construir un modelo predictivo basado en dos enfoques: el primero, utilizando modelos de econometría clásica fundamentados en el análisis de series de tiempo, y el segundo, mediante el aprendizaje de máquina utilizando modelos de aprendizaje profundo. Estas fases se desarrollan siguiendo la metodología DST (Trayectoria de Ciencia de Datos), que integra las actividades de CRISP-DM, sin seguir un orden predeterminado. En cambio, se secuencian de acuerdo con la información disponible y los resultados de las fases previas.

El estudio buscó responder a las siguientes preguntas: ¿Cuáles variables exógenas contribuyen a la predicción del AUM (Activos Bajo Administración) de los Fondos de Inversión Colectiva? y ¿Qué modelo lograba captar mejor las relaciones de las series de tiempo que conforman dichas variables?

Palabras claves: Inteligencia Artificial, Series Temporales, Mercado Financiero.

Abstract

With an average of \$72 billion under management over the past five years, mutual funds have become a preeminent investment vehicle. These funds have shown a compound annual growth rate of 6.35%, with the number of investors increasing by an average 10% per year. Despite positive indicators and high returns, no modeling studies have been developed to predict the behavior of assets under management. Such a forecast could help anticipate periods of liquidity shortages.

This project aimed to develop a predictive model using two approaches: classical econometric models and machine learning. These phases are developed according to the DST methodology, which integrates CRISP-DM activities without following a predetermined order. Instead, they are sequenced according to the available information and the results of the previous phases.

The project sought to answer the following questions ¿Which exogenous variables contribute to the prediction of the AUM (Assets Under Management) of mutual funds? and ¿Which model was able to better capture the relationships of the time series that make up these variables?

Keywords: Artificial Intelligence, Time series, Financial Markets.

Tabla de Contenido

Introducción	13
Justificación	15
Objetivos.....	19
Objetivo General	19
Objetivos Específicos.....	19
Descripción del Problema	20
Marco Conceptual.....	22
Fondos de Inversión Colectiva.....	22
Series de Tiempo	26
Modelos Autorregresivos	27
Modelos ARIMAX	28
Modelos VAR	28
Modelos de Suavizado Exponencial.....	30
Suavizado Exponencial Simple.	30
Suavizamiento Exponencial Lineal De Holt.	31
Suavizamiento Exponencial de Tendencia Amortiguada.	31
Suavizamiento Exponencial de Winters Aditivo.	31
Suavizamiento Exponencial de Winters Multiplicativo.	32
Redes LSTM.....	32
Compuertas.....	33
Actualización de la Celda de Estado.	34
Nuevo Estado Oculto.....	34

Gradient Boosting.....	35
XGBoost.....	36
LightGBM.....	37
Prophet.....	37
Estado del Arte.....	37
Metodología.....	41
Método.....	41
Procedimiento.....	43
Desarrollo.....	46
Entendimiento del Negocio.....	46
Exploración de Objetivos.....	49
Exploración Fuentes de Datos y Extracción de Datos.....	51
Entendimiento de los Datos.....	55
Preparación de Datos.....	58
Caracterizaciones Variables Económicas.....	65
Modelado de Datos.....	70
Modelo ARIMA.....	71
Modelo ARIMAX.....	73
Modelo VAR.....	75
Suavizado Exponencial.....	77
Modelo LSTM.....	80
LightGBM.....	83
Prophet.....	86

Resultados	88
Resultados de la Evaluación	88
Análisis de Resultados	90
Evaluación del Desempeño del Modelo LSTM Multivariado	92
Conclusiones	94
Recomendaciones	96
Referencias.....	98
Apéndices.....	108

Lista de Tablas

Tabla 1 <i>Posibles Variables Predictoras</i>	47
Tabla 2 <i>Descripción Conjuntos de Datos</i>	55
Tabla 3 <i>Variables Seleccionadas Después del Filtrado de Variables</i>	57
Tabla 4 <i>Resultados Prueba de Anderson - Darling</i>	66
Tabla 5 <i>Resultados Modelo ARIMA</i>	72
Tabla 6 <i>Resultados Modelo ARIMAX</i>	74
Tabla 7 <i>Configuración Hiperparámetros Modelos LSTM</i>	82
Tabla 8 <i>Comparación Métricas de Evaluación Modelos Predictivos AUM FIC</i>	89

Lista de Figuras

Figura 1 <i>Distribución Porcentual AUM por Tipo de Fondo</i>	15
Figura 2 <i>Distribución Porcentual Inversionistas por Tipo de FIC</i>	16
Figura 3 <i>Variación AUM por Componentes Año 2024</i>	16
Figura 4 <i>Celda LSTM</i>	33
Figura 5 <i>Secuencia Red LSTM</i>	35
Figura 6 <i>Proceso de Boosting</i>	36
Figura 7 <i>Flujograma Metodología DST</i>	45
Figura 8 <i>Diagrama de Árbol de Objetivos para la de Intervención del Problema</i>	51
Figura 9 <i>Fases Ejecución Estudio</i>	51
Figura 10 <i>Resumen Fuentes de Datos</i>	54
Figura 11 <i>Identificación Gráfica de Valores Inconsistentes</i>	60
Figura 12 <i>Serie Histórica AUM FIC Fiduciarias</i>	62
Figura 13 <i>Aportes Recibidos y Retiros por Categoría de FIC</i>	64
Figura 14 <i>Actividades de Preprocesamiento de Datos</i>	64
Figura 15 <i>Matriz de Correlación de Variables</i>	69
Figura 16 <i>Diagrama de Flujo Procedimiento Modelo ARIMA</i>	71
Figura 17 <i>Gráfico AUM Dato Real vs Predicción ARIMA</i>	72
Figura 18 <i>Diagrama de Flujo Procedimiento Modelo ARIMAX</i>	73
Figura 19 <i>Gráfico AUM Dato Real vs Predicción ARIMAX</i>	74
Figura 20 <i>Diagrama de Flujo Procedimiento Modelo VAR</i>	76
Figura 21 <i>Gráfico AUM Dato Real vs Predicción VAR</i>	77
Figura 22 <i>Descomposición Multiplicativa de la Serie por Año</i>	78

Figura 23 <i>Diagrama de Flujo Procedimiento Modelo Suavizado Exponencial</i>	79
Figura 24 <i>Gráfico AUM Dato Real vs Predicción Suavizado Exponencial</i>	80
Figura 25 <i>Diagrama de Flujo Procedimiento Modelo LSTM</i>	81
Figura 26 <i>Gráfico AUM Dato Real vs Predicción LSTM</i>	82
Figura 27 <i>Diagrama de Flujo Procedimiento Modelo LigthGBM</i>	84
Figura 28 <i>Gráfico AUM Dato Real vs Predicción LightGBM</i>	85
Figura 29 <i>Diagrama de Flujo Procedimiento Determinación Modelo Prophet</i>	86
Figura 30 <i>Gráfico AUM Dato Real vs Predicción Prophet</i>	87
Figura 31 <i>Tendencias Captadas por el Modelo Prophet Multivariado</i>	87
Figura 32 <i>Composición del Portafolio de Inversión de los FIC</i>	91
Figura 33 <i>Gráfico AUM Dato Real vs Predicción Media Móvil</i>	93

Lista de Apéndices

Apéndice A <i>Características Conjuntos de Datos AUM FIC</i>	108
Apéndice B <i>Características Conjunto de Datos Agregados Monetarios Crediticios</i>	109
Apéndice C <i>Características Conjunto de Datos Índice de Precios al Consumidor (IPC)</i>	113
Apéndice D <i>Características Conjunto de Datos Tasa de Intervención de Política Monetaria</i> ..	114
Apéndice E <i>Características Conjunto de Datos Categorización FIC</i>	115
Apéndice F <i>Características Conjunto de Datos Tasas de Captación CDT</i>	116
Apéndice G <i>Características Conjunto de Datos Valor COLTES</i>	119
Apéndice H <i>Características Conjunto de Datos Valor Dólar</i>	120

Introducción

Los Fondos de Inversión Colectiva (FIC) han emergido como uno de los principales vehículos de inversión en el sistema financiero colombiano, alcanzando un crecimiento significativo en los últimos años. Con un aumento del 133% en sus activos administrados durante la última década y una participación del 5,4% el PIB al cierre de 2023 (Asociación de fiduciarias de Colombia, 2024), estos fondos han demostrado ser una opción atractiva tanto para inversionistas institucionales como para personas naturales y jurídicas. No obstante, a pesar de su relevancia en el mercado financiero, el sector fiduciario aún carece de herramientas prospectivas que permitan prever con mayor precisión el comportamiento de estos activos administrados, lo que limita la capacidad de anticipación y mitigación de riesgos financieros.

El presente estudio aborda esta problemática mediante la implementación de modelos predictivos basados en inteligencia artificial y técnicas econométricas, con el objetivo de desarrollar un marco analítico que optimice la toma de decisiones en la gestión de los FIC. Para ello, se emplea una metodología que integra modelos de series temporales tradicionales, como ARIMA y VAR, con enfoques avanzados de aprendizaje automático, como redes neuronales LSTM y algoritmos de boosting. La combinación de estas técnicas permite capturar patrones complejos en los datos y mejorar la precisión en la estimación del comportamiento de los activos bajo administración.

Además, el estudio busca identificar y caracterizar las variables exógenas con mayor incidencia en la evolución de los FIC, considerando factores macroeconómicos como tasas de interés, agregados monetarios y movimientos en el mercado de deuda. A través de un análisis riguroso de correlación estadística y causalidad, se pretende construir un modelo que no solo

prediga el comportamiento de los activos, sino que también aporte información valiosa sobre las dinámicas subyacentes que afectan su evolución.

La investigación se estructura en diversas fases, iniciando con el entendimiento del negocio y la exploración de objetivos, seguido por la adquisición y preparación de datos, y culminando con la implementación y validación de modelos.

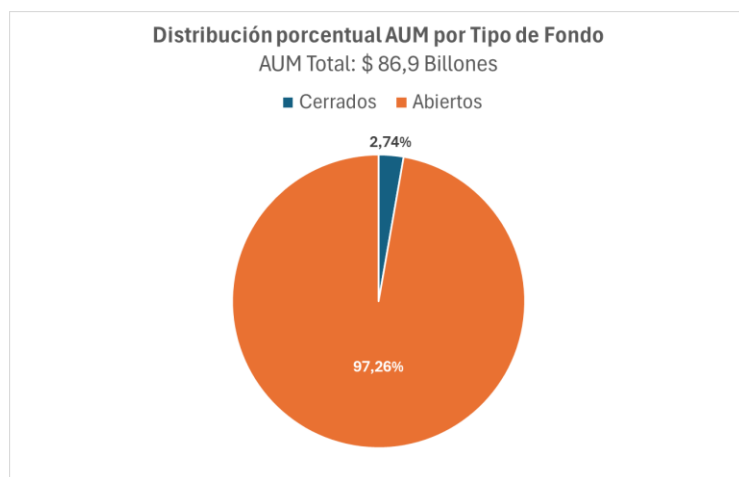
En este contexto, el estudio pretende llenar un vacío existente en la literatura y en la industria financiera colombiana, proporcionando herramientas basadas en ciencia de datos para mejorar la planificación estratégica y la gestión de riesgos en los Fondos de Inversión Colectiva. Con ello, se espera contribuir al fortalecimiento del sector fiduciario y a la optimización de sus procesos de inversión y administración de activos.

Justificación

El presente trabajo de grado se enfoca en la necesidad que tiene el sector fiduciario colombiano, compuesto por 27 sociedades fiduciarias, agremiadas a través de la Asociación de Fiduciarias (que agremia 25 de estas 27 entidades), de contar con herramientas predictivas para prever con cierto grado de precisión el comportamiento de los activos administrados por los fondos de inversión colectiva clasificados como fondos abiertos o fondos vista. ¿De dónde radica esta necesidad? Los fondos de inversión colectiva abiertos, al tener la naturaleza de fondos “en los cuales la Sociedad Administradora está obligada a redimir las participaciones de los inversionistas en cualquier momento, durante la vigencia de este” (Decreto 2555, 2010, art.3.11.2.3) son la tipología de fondos que más AUM e inversionistas agrupan en toda la industria de fondos fiduciarios como lo muestran la Figura 1 y la Figura 2.

Figura 1

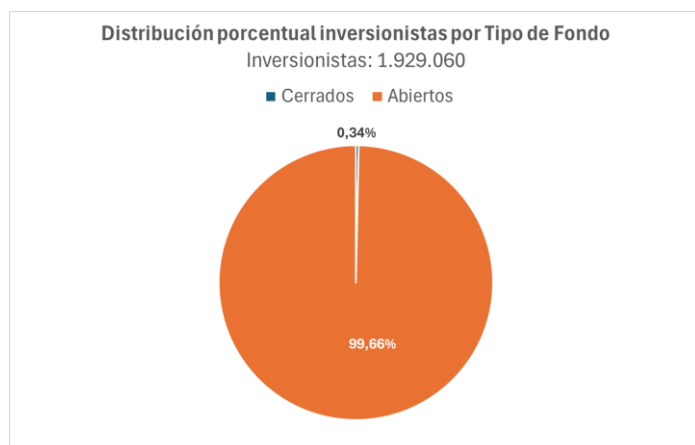
Distribución Porcentual AUM por Tipo de Fondo



Nota. Cifras con fecha de corte 30 de abril de 2024, no incluyen Fondos de Capital Privado.

Figura 2

Distribución Porcentual Inversionistas por Tipo de FIC

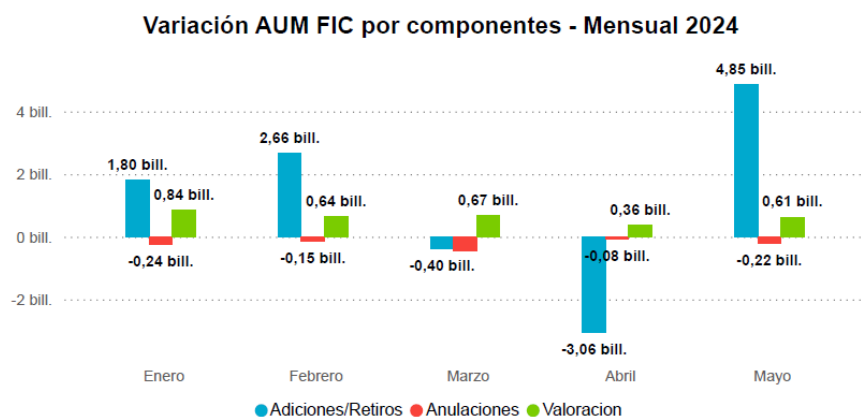


Nota. Cifras con fecha de corte 30 de abril de 2024, no incluyen Fondos de Capital Privado.

Por otro lado, si se observa cual ha sido el comportamiento de los componentes de la dinámica de los FIC – Adiciones/Retiros, Anulaciones y Valoración – se puede observar que el año 2024 a corte de mayo se ha caracterizado por ser un año en el que el aumento del AUM se explica principalmente por \$ 5.85 billones de adiciones netas, \$ 3.12 billones de valoración de las inversiones y \$ 1.14 billones de transacciones anuladas, como lo muestra la Figura 3.

Figura 3

Variación AUM por Componentes Año 2024



De aquí se puede concluir que el año tiene una tendencia positiva de crecimiento de los activos administrados, explicada por una cantidad de aportes que supera en un 87 % a la valoración de las inversiones, que, en otros términos, indica que los inversionistas sustentan el crecimiento del AUM con el ingreso de una mayor cantidad de activos. Sin embargo, al fijar la mirada en un mes como abril de 2024, se observa que este periodo estuvo marcado por retiros de los mismos inversionistas por el orden de \$3.06 billones y únicamente \$ 0.36 billones de valoración.

Aquí surgen varias preguntas ¿Qué causa que el año corrido se enmarque en un periodo de crecimiento constante? ¿Cuáles son las variables de tipo económico o financiero que afectaron el mes de abril? ¿Existe estacionalidad en el ritmo del AUM de FIC abiertos? ¿Las variables macroeconómicas tienen una incidencia fuerte en las adiciones o retiros, o por el contrario son variables de tipo específico? Estos interrogantes son los que la industria fiduciaria en el país se ha planteado y a los cuáles les busca respuesta, puesto que al tener, de mínimo una aproximación, ayudaría en gran medida a entender el mercado y posteriormente a prever circunstancias adversas como las de un mes con las características de abril de 2024 en la que la salida de recursos en una magnitud considerable podría afectar la liquidez de los fondos, la venta de inversiones con carácter de urgencia a un menor precio para cumplir con los compromisos adquiridos con los inversionistas y otra serie de efectos adversos en las inversiones a los cuales se les podría anticipar con estrategias adecuadas.

Es así como, a través de este estudio, se pretendió analizar la serie histórica de AUM de forma individual y en conjunto con otras series históricas de variables exógenas que se “presumía” son las más adecuadas para explicar el comportamiento de los activos en los FIC, de forma que se sienten las bases para plantear un modelo predictivo que traiga a la discusión del

sector un enfoque prospectivo que enriquezca los esfuerzos individuales de cada fiduciaria y de esta manera alinear estrategias con las variables observadas y mitigar algún tipo de crisis que pueda derivar de un desconocimiento de la relación de la misma dinámica del mercado con el entorno.

Objetivos

Objetivo General

Identificar el mejor modelo predictivo para los activos administrados por las fiduciarías en los Fondos de Inversión Colectiva en Colombia utilizando herramientas econométricas y de aprendizaje automático

Objetivos Específicos

Identificar las variables económicas que influyen en el comportamiento de los activos administrados por las Sociedades Fiduciarías a través de los Fondos de Inversión Colectiva en Colombia mediante correlación estadística.

Implementar técnicas econométricas y aprendizaje automático para estimar el comportamiento de los activos administrados por las Sociedades Fiduciarías a través de los Fondos de Inversión Colectiva en Colombia.

Validar el desempeño de los modelos implementados a través de las métricas de evaluación RMSE (Root Mean Square Error) y MAE (Mean Absolute Error) para identificar el modelo que mejor describa el comportamiento de los activos administrados por los Fondos de Inversión Colectiva administrados por las Sociedades Fiduciarías.

Descripción del Problema

Los Fondos de Inversión Colectiva (FIC) en Colombia han tenido un ascenso gradual en términos de sus activos administrados (AUM) en los últimos 10 años, con una variación de 133%, ubicándolos como un instrumento de inversión que representa un 5,4% del porcentaje del PIB (Producto Interno Bruto) al cierre del año 2023 y que además lo deja en un tercer lugar si se compara con otros productos de inversión y agregados monetarios. Perspectiva que cobra relevancia cuando se observa una variación de 83% para los últimos 10 años en el número de inversionistas, ubicándose al cierre de 2023 en más de dos millones de clientes – tanto personas naturales como jurídicas – que confían sus recursos a este vehículo de inversión (Asociación de fiduciarias de Colombia, 2024).

Sin embargo, a pesar de estas cifras que reflejan una evolución positiva en la generalidad de los fondos, su comportamiento no se ha modelado a través de una visión prospectiva, al menos desde el gremio, sumado a que en un ejercicio de análisis estratégico llevado a cabo por el sector, en cabeza de la Asociación de Fiduciarias, se listaron una serie de temas materiales que se debían atender con un nivel de prioridad media, ya que se encontraron brechas en su gestión, entre los que destaca la Innovación y Transformación Digital (Asofiduciarias, 2022 , p. 47).

La ausencia de información que permita generar conocimiento prospectivo a nivel gremial para la gestión óptima de los FIC en el sector fiduciario constituye una deficiencia histórica. Esto se debe, en parte, a la limitada aplicación de nuevas tendencias en gestión de datos fundamentadas en ciencias computacionales, un desafío crítico frente a las crecientes demandas globales de procesamiento avanzado de información (Mejía y Mejía, 2020, p. 59). Estas limitaciones se evidencian en la falta de adopción efectiva de modelos de aprendizaje

automático para los FIC, como los planteados por Vergara Cardona (2022) y Salgado Gaspar (2024), los cuales no han sido integrados con suficiente contundencia.

Por otro lado, este panorama está directamente relacionado con el desconocimiento o la imprecisión en la identificación de las variables clave que influyen en la gestión de los fondos y su impacto, que responden a fuerzas externas, representadas por cambios en un conjunto de variables macroeconómicas (Das et al., 2005; Rayón, 2015). Como consecuencia, emergen deficiencias significativas en la gestión de datos, lo que no solo puede resultar en estrategias desalineadas con las perspectivas económicas, sino también en un aumento del riesgo de crisis de liquidez debido a retiros inesperados y a su vez, teniendo en cuenta que cada vez son más las personas que depositan su confianza en los FIC, una afectación de este tipo puede afectar al funcionamiento y estabilidad del sistema financiero. Dichos eventos podrían mitigarse o, al menos, gestionarse de manera más eficiente mediante un enfoque prospectivo basado en herramientas que permitan evidenciar y anticipar vulnerabilidades de los FIC (Rendón-González, 2021).

En conclusión, una apropiada gestión de datos en relación con la aplicación de modelos de aprendizaje automático, que se conviertan, como menciona (Rendón-González, 2021) en una línea de investigación con mayores componentes cuantitativos y alineada con las necesidades de cuantificación de riesgos e identificación de vulnerabilidades de los FIC, se basa en la evaluación del efecto de variables macroeconómicas sobre el comportamiento de dichos fondos. Por lo tanto, teniendo en cuenta la problemática anterior, se propone la siguiente pregunta de investigación ¿Cómo pronosticar el comportamiento de los activos administrados por los Fondos de Inversión Colectiva utilizando herramientas de inteligencia artificial?

Marco Conceptual

Fondos de Inversión Colectiva

Los Fondos de Inversión Colectiva (o también llamados FIC) que tienen a nivel internacional sus pares conocidos como Fondos Mutuos, en el contexto colombiano son definidos según el Decreto 1242 de 2013 como:

Todo mecanismo o vehículo de captación o administración de sumas de dinero u otros activos, integrado con el aporte de un número plural de personas determinables una vez el fondo entre en operación, recursos que serán gestionados de manera colectiva para obtener resultados económicos también colectivos. (Decreto 1242, 2013, Artículo 3.1.1.2.1)

Es decir, son un producto que una entidad especializada gestiona y administra para captar recursos ya sea monetarios o en forma de activos que aportan un número de inversionistas para así generar rendimientos según los reglamentos establecidos, para proporcionalmente a la inversión de cada participante distribuir los rendimientos obtenidos (Mejía y Mejía, 2020, p. 7).

Esta redención de rendimientos depende del tipo de FIC y lo definido en su reglamento y dependen de varios factores tales como las comisiones cobradas por la sociedad administradora, tiempo de permanencia de la inversión, nivel de riesgo y variables del mercado (Ortega y Quirama, 2020, p.21), teniendo en cuenta que el riesgo de pérdida por las variaciones en el precio de los activos que forman parte de la inversión es asumido por el inversionista (Ramírez, 2012, p.16) y la sociedad administradora solo es responsable de tomar las decisiones de “manera profesional, con la diligencia exigible a un experto prudente y diligente en la gestión de fondos de inversión colectiva de inversión del fondo de inversión colectiva, el reglamento y las normas aplicables” (Decreto 1242, 2013, Artículo 3.1.5.2.1), de tal manera que se puede decir que la

sociedad administradora gestiona un portafolio con obligaciones de medio y no de resultado, ya que los dineros son entregados a títulos de inversión y no de depósito, lo que no genera la obligación que tienen las instituciones de depósito. Es así como la sociedad fiduciaria tiene como primera obligación entre otras, favorecer en todo caso los intereses de los inversionistas (Decreto 1242, 2013, Artículo 3.1.3.1.3)

Dentro de los tipos de Fondos se enmarcan dos grandes categorías como los son los Fondos de Inversión colectiva abiertos “en los cuales la Sociedad Administradora está obligada a redimir las participaciones de los inversionistas en cualquier momento, durante la vigencia del mismo” (Decreto 1242, 2013, Artículo 3.1.1.2.3) y los Fondos de Inversión Colectiva Cerrados que:

son aquellos Fondos de Inversión Colectiva en los que la Sociedad Administradora de Fondos de Inversión Colectiva únicamente está obligada a redimir las participaciones de los inversionistas al final del plazo previsto para la duración del Fondo de Inversión Colectiva. (Decreto 1242, 2013, Artículo 3.1.1.2.3)

Cabe anotar que, con relación al plan de inversión, se clasifican diferentes tipos de FICs de acuerdo con cómo se destinen los recursos en el marco de una política de inversión: FIC de mercado monetario, FIC inmobiliario, FIC Bursátil, y Fondos de Capital Privado (Asociación Nacional de Instituciones Financieras [ANIF], 2016, p. 52).

Los FIC del mercado monetario son:

aquellos Fondos de Inversión Colectiva abiertos sin pacto de permanencia, cuyo portafolio esté constituido exclusivamente con valores denominados en moneda nacional o unidades representativas de moneda nacional, inscritos en el Registro Nacional de Valores y Emisores (RNVE), calificados por una sociedad legalmente habilitada para el

efecto con mínimo de grado de inversión, salvo los títulos de deuda pública emitidos o garantizados por la Nación, por el Banco de la República o por el Fondo de Garantías de Instituciones Financieras (Fogafín), los cuales no requerirán calificación. (Decreto 1242, 2013, Artículo 3.2.1.1.1)

Los Fondos de Capital Privado son:

Fondos de Inversión Colectiva cerrados que deben destinar al menos las dos terceras partes (2/3) de los aportes de sus inversionistas a la adquisición de activos o derechos de contenido económico diferentes a valores inscritos en el Registro Nacional de Valores y Emisores (RNVE). (Decreto 1242, 2013, Artículo 3.3.1.1.2).

Por otro lado, los Fondos Bursátiles son los “Fondos de Inversión Colectiva cuyo objeto consista en replicar o seguir un índice nacional o internacional, mediante la constitución de un portafolio integrado por algunos o todos los activos que hagan parte de la canasta que conforma dicho índice” (Decreto 1242, 2013, Artículo 3.4.1.1.2).

Y por último se tienen los FIC Inmobiliarios que son los “Fondos de Inversión Colectiva cerrados cuya política de inversión prevea una concentración mínima en los activos mencionados, equivalente al setenta y cinco por ciento (75%) de los activos totales del fondo” (Decreto 1242, 2013, Artículo 3.5.1.1.1). Este tipo de activos son bienes inmuebles tanto en Colombia como en el exterior, títulos emitidos dentro de procesos de titularización hipotecaria o inmobiliaria, derechos fiduciarios de patrimonios autónomos conformados por bienes inmuebles que tengan como fin un proyecto inmobiliario, participaciones en otros FIC inmobiliarios locales, proyectos inmobiliarios en Colombia o el exterior o acciones o cuotas de interés social que emitan compañías nacionales o extranjeras que tengan por objeto exclusivo la inversión en bienes inmuebles, proyectos inmobiliarios (Decreto 1242, 2013, Artículo 3.5.1.1.2).

Es importante mencionar que los FIC actualmente pueden ser constituidos únicamente por Sociedades Comisionistas de Bolsa (SCB) y por Sociedades Fiduciarias (SF) las cuales tienen actividades claramente definidas dentro del funcionamiento de los FIC como lo son la administración (manejo administrativo hasta la contratación y entrega en custodia de los valores), la gestión (toma de decisiones de las inversiones), la custodia (cuidado y vigilancia de los activos para cumplir las operaciones de mercado) y la distribución (promoción para vincular inversionistas a los FIC) (ANIF, 2016, p. 54-55) .

Dentro del mercado de los FIC en Colombia, las sociedades fiduciarias continúan siendo los principales administradores de los recursos de los inversionistas, concentrando un 72,32% de participación en el mercado a corte de abril de 2024. Aunque esta cifra ha mostrado una tendencia a la baja en comparación con el mismo periodo hace cinco años, las sociedades fiduciarias siguen desempeñando un papel preponderante en la industria de FIC y Fondos de Capital Privado (FCP) en el país.

En este contexto, el análisis predictivo se ha convertido en una herramienta esencial para el sector fiduciario, impulsado por la creciente necesidad de optimizar la gestión de recursos y mejorar la toma de decisiones mediante el uso de modelos analíticos avanzados. La aplicación de estos modelos, junto con la implementación de herramientas de tecnología e inteligencia artificial, permite generar proyecciones más precisas y oportunas sobre la evolución de los activos administrados.

Uno de los principales desafíos del sector es identificar qué variables determinan el monto de recursos administrados por los FIC y evaluar si, a través de modelos predictivos, es posible estimar el comportamiento futuro de estos activos en función del análisis histórico del AUM administrado por las sociedades fiduciarias. Este enfoque no solo fortalecería la

planificación estratégica y la gestión del riesgo, sino que también consolidaría el papel de la analítica avanzada en la evolución del mercado fiduciario en Colombia.

Series de Tiempo

Una serie de tiempo se podría definir como una “serie de puntos tomada secuencialmente a lo largo del tiempo o como el resultado de un proceso estocástico” (Thanaki, 2018, p. 2) o como la define (Pal y Prakash, 2017) una “composición de observaciones cuantitativas sobre una o más características medibles de una entidad individual y tomadas en múltiples puntos en el tiempo” (p.16) que a su vez poseen unas estructuras de tipo interno que la caracterizan, a decir: tendencia, estacionalidad, estacionariedad, autocorrelación, entre otras.

La tendencia general de una serie de tiempo se puede definir como movimientos a largo plazo de los valores o en términos más prácticos como un “aumento o disminución a largo plazo de una secuencia” (Thanaki, 2018, p. 4) y puede ser modelada como una regresión de la serie contra el tiempo y otras variables que se conocen como variables explicativas (Pal y Prakash, 2017, p.23). Estas tendencias se usan para prever el futuro con respecto a las observaciones del pasado, proceso conocido como forecasting, y que a su vez puede tomar características adicionales, llamadas características exógenas (Thanaki, 2018, p. 3).

La estacionalidad se puede definir como variaciones repetitivas y periódicas en una serie temporal (Pal y Prakash, 2017, p. 26; Thanaki, 2018, p. 5) el periodo en el que se observa la estacionalidad debe ser menor de un año, tales como, periodos diarios, semanales, mensuales o anuales. De igual manera hay “periodos cíclicos que ocurren con menos frecuencia que las fluctuaciones estacionales, sin tener periodos fijos de ocurrencia” (Pal y Prakash, 2017, p. 31) y otros periodos que no son cíclicos y que explican la variabilidad, que se podrían denominar

irregulares, como decisiones de negocio, campañas publicitarias, la introducción de un nuevo producto, etc. (Thanaki, 2018, p. 5).

Aunque las variaciones de la temporalidad se pueden explicar por la tendencia, la estacionalidad, los cambios cíclicos y no cíclicos, existe un componente que:

Refleja variaciones inesperadas en la serie temporal. Las variaciones inesperadas son estocásticas y no pueden enmarcarse en un modelo matemático para una predicción futura definitiva. Este tipo de error se debe a la falta de información sobre variables explicativas que puedan modelizar estas variaciones o a la presencia de un ruido aleatorio. (Pal y Prakash, 2017, p. 32)

Modelos Autorregresivos

Los modelos autorregresivos son los que usan términos de retardo para hacer regresión sobre series temporales, es decir, predicen valores futuros basándose en valores temporales anteriores autorregresivo (Pal y Prakash, 2017, pp. 114). Box y Jenkins desarrollaron modelos estadísticos para series temporales que tienen en cuenta la dependencia existente entre estos datos. Estos modelos se conocen con el nombre de ARIMA (AutoRegresive Integrated Moving Average) derivados de sus tres componentes AR (Autoregresivo), I (Integrado) y MA (Medias Móviles) (De la Fuente Fernández, s.f.).

El modelo se representa por el orden de los componentes AR, MA y de la diferenciación

$$y_t^{(d)} = c + \varepsilon_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

En la Ecuación 1 el componente de diferenciación se encuentra en verde, el componente AR en azul y el componente MA en rojo.

Para aplicar los modelos ARIMA se puede seguir la siguiente metodología aplicada en 4 fases según lo descrito por (De la Fuente Fernández, s.f.):

- Identificar el posible modelo ARIMA que sigue la serie, lo que incluye decidir qué transformaciones usar para convertir la serie en una serie estacionaria y determinar los órdenes p y q de su estructura autorregresiva y de media móvil;
- Los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen los errores estándar y los residuos del modelo;
- Se comprueba que los residuos no tienen una estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran una estructura se modifica el modelo y se repiten las etapas anteriores hasta obtener un modelo adecuado;
- Por último, una vez se obtuvo el modelo se realizan las predicciones con este. (p 1).

Modelos ARIMAX

El modelo ARIMAX es el modelo autorregresivo que incluye variables exógenas, definidas como las variables que vienen dadas fuera del modelo, por tal razón, es conveniente estudiar las trayectorias temporales de las variables endógenas, asociadas con posibles trayectorias temporales alternativas de las variables exógenas (Sargent, 1988). La Ecuación 2 muestra la forma que toma un modelo ARIMAX

$$y_t^{(d)} = c + \varepsilon_t + \beta X + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

Como se puede observar se trata de la misma ecuación de un modelo ARIMA, adicionando una variable adicional (variable exógena)

Modelos VAR

Un VAR (Vector Autorregresivo) es un modelo que involucra ecuaciones simultáneas formando un sistema de ecuaciones en los cuales los valores contemporáneos de las variables del modelo no aparecen como variables explicativas en ninguna de las ecuaciones. En sentido

contrario, las variables explicativas de cada una de las ecuaciones es el conjunto del bloque de retardos de cada una de las variables del modelo (Novales, 2017, p. 2). En otras palabras, cada variable es explicada por los rezagos temporales de las demás variables del sistema. Esta técnica se usa para predecir series de tiempo que se interrelacionan. El modelo parte de un sistema estructura dinámico como se muestra en la Ecuación 3.

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + Y_1 z_t + \epsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + Y_2 z_t + \epsilon_{2t} \end{aligned} \quad (3)$$

El sistema de ecuaciones simultáneas tiene dos variables endógenas y_{1t} y y_{2t} las cuales son estacionarias, ϵ_{1t} y ϵ_{2t} son procesos de ruido blanco y z_t es un vector variables exógenas. En este sistema solo se incluye un retardo para efectos explicativos, sin embargo, identificar el número de retardos hace parte del análisis que se debe realizar para especificar el modelo (Novales, 2017, p. 3).

Las siguientes son algunas consideraciones sobre la estimación de un modelo VAR según (Novales, 2017):

- En un modelo VAR todas las variables son tratadas simétricamente, siendo explicadas por los valores pasados de todas ellas. Teniendo el modelo tantas ecuaciones como variables, y los valores de retardo de todas las ecuaciones aparecen como variables explicativas en todas las ecuaciones;
- Cuando se estima el modelo, se pueden excluir algunas variables explicativas, en función de su significación estadística;
- En el modelo VAR se pueden estimar con bastante precisión los elementos globales del modelo, entre estos, el R2 y el efecto global de una variable sobre otra, lo que lleva a contrastes de causalidad, como el de causalidad en el sentido de Granger. (p. 7)

Modelos de Suavizado Exponencial

Debido a que las series temporales tienen un componente de señales que captan la dinámica del proceso y un componente de ruido que representa el componente no modelado del proceso, el enfoque basado en suavizado se basa en filtros que toman las entradas y las separan en componentes de tendencia y de ruido (Pal y Prakash, 2017, p. 87). Es así como el suavizado elimina el componente residual o de ruido, capturando los componentes de tendencia y estacionalidad. Dentro de estos modelos resaltan los de Suavizado Exponencial Simple, Suavizamiento Exponencial Lineal de Holt, Suavizamiento Exponencial de Tendencia Amortiguada, Suavizamiento Exponencial de Winters Aditivo, Suavizamiento Exponencial de Winters Multiplicativo.

Suavizado Exponencial Simple. Este modelo se basa en que el valor futuro de una serie temporal se calcula como la combinación de los valores previos, asignando pesos exponenciales decrecientes a estos valores, teniendo a los valores más recientes con un impacto mayor en la predicción que los valores anteriores. Se usa para datos con tendencias, pero sin patrones de estacionalidad clara (Mora Caballero, 2023). El cálculo se realiza mediante la Ecuación 4.

$$Y_{t+1} = \alpha X_t + (1 - \alpha) Y_t \quad (4)$$

Donde:

Y_{t+1} : Pronóstico para un periodo futuro

α : constante de suavización, entre 0 y 1

X_t : Valor real para el periodo de tiempo

Y_t : Pronóstico hecho previamente para el periodo de tiempo

Suavizamiento Exponencial Lineal De Holt. Se considera una extensión del suavizamiento exponencial simple que incluye un componente para la captura de tendencia lineales en los datos. Se usa cuando los datos muestran una tendencia a largo plazo, además de una variación estacional. Para suavizar los datos el modelo usa dos parámetros de suavizamiento, α (alfa) para suavizar el nivel y β (beta) para suavizar la tendencia (Mora Caballero, 2023) según la Ecuación 5.

$$F_{t+p} = L_t + pb_t$$

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (5)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)(b_{t-1})$$

Suavizamiento Exponencial de Tendencia Amortiguada. También es una extensión del modelo de suavizado exponencial lineal de Holt, pero a diferencia de este no supone una tendencia constante, sino que permite que la tendencia disminuya gradualmente con el tiempo de manera no lineal, como lo muestra la Ecuación 6. El factor de amortiguación controla la tasa de disminución de la tendencia, es un número entre 0 y 1. Un valor de 1 muestra una tendencia que cambia lentamente y un valor de 0 que la tendencia cambia rápidamente (Mora Caballero, 2023).

$$F_{t+p} = L_t + pb_t$$

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + \varphi b_{t-1}) \quad (6)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)(\varphi b_{t-1})$$

Suavizamiento Exponencial de Winters Aditivo. Es un modelo que se usa cuando los datos muestran patrones estacionales, agrega un componente para modelar la estacionalidad de manera más efectiva, como se muestra en la Ecuación 7, útil para series de tiempo que exhiben fluctuaciones regulares a lo largo del tiempo (Mora Caballero, 2023).

$$F_{t+p} = L_t + pb_t$$

$$L_t = \alpha(Y_t - sn_{t-1}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (7)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)(b_{t-1})$$

$$s_{nt} = \delta(Y_t - L_t) + (1 - \delta)s_{n-1}$$

Suavizamiento Exponencial de Winters Multiplicativo. Es un modelo que se usa cuando los datos muestran patrones estacionales y la magnitud de la estacionalidad es proporcional a la magnitud de la tendencia como lo muestra la Ecuación 8 (Mora Caballero, 2023).

$$F_{t+p} = L_t + pb_t$$

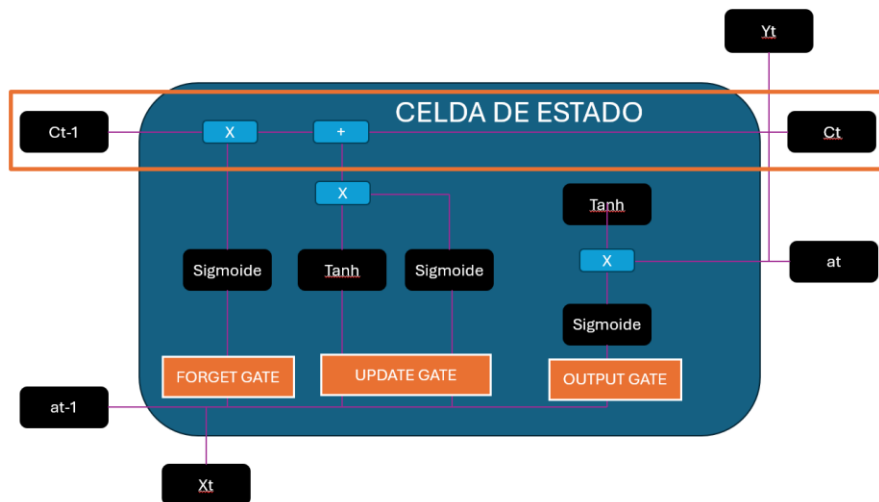
$$L_t = \alpha(Y_t / sn_{t-1}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)(b_{t-1})$$

$$s_{nt} = \delta(Y_t/L_t) + (1 - \delta)s_{n-1}$$

Redes LSTM

La célula o red LSTM (Long Short-Term Memory) es un tipo de célula RNN (Recurrent Neuronal Net) que añade memoria a largo plazo de una forma eficaz permitiendo aprender aún más parámetros, como se observa en la Figura 4. Lo que la convierte en la RNN más potente para hacer pronósticos cuando se tienen tendencias a largo plazo en los datos (Korstanje, 2021, p. 243).

Figura 4*Celda LSTM*

Nota. En esta red se pueden ver dos pesos c y a del tiempo $t-1$ y luego de ser transformados c y a del tiempo t . La X es la entrada a esta celda y la Y es la salida.

Compuertas. La celda de estado es donde se pueden añadir o remover datos que no son útiles para la red. Las compuertas son los mecanismos usados para la adición o remoción de datos:

1. La forget gate, permite que los elementos de memoria sean eliminados
2. La update gate, permite añadir nuevos elementos a la memoria
3. La output gate, permite la creación del estado oculto actualizado

Cada compuerta se encuentra conformada por una red neuronal, una función sigmooidal y un elemento multiplicador. La función sigmooidal es la que se encarga de “decidir” que información se va a eliminar o descartar, puesto que, si alcanza valores cercanos a cero, anula los datos de entrada y si son valores cercanos a 1, permite el paso de estos (Codificando Bits, 2019).

Para el caso específico de la forget gate toma el estado oculto anterior a_{t-1} y la entrada actual X_t , los transforma y los lleva a la función de activación sigmoial, generando el vector F_t como lo muestra la Ecuación 9 (Codificando Bits, 2019).

$$f_t = \text{sigmoidal} (W_f [a_{t-1}, X_t] + b_f) \quad (9)$$

Si alguno de estos valores de este vector es 0 o muy cercano a 0, la red eliminará esta información, pero si son cercanos a 1, los enviará a la celda de estado, alimentando la memoria (Codificando Bits, 2019). El funcionamiento de la update gate es similar: se toma el estado oculto anterior a_{t-1} y la entrada actual X_t , los transforma y los lleva a la función de activación sigmoial, generando el vector U_t , como lo muestra la Ecuación 10 (Codificando Bits, 2019).

$$u_t = \text{sigmoidal} (W_i [a_{t-1}, X_t] + b_i) \quad (10)$$

Los valores cercanos a 1 de este vector son los que se quieren conservar en la memoria de la red (Codificando Bits, 2019).

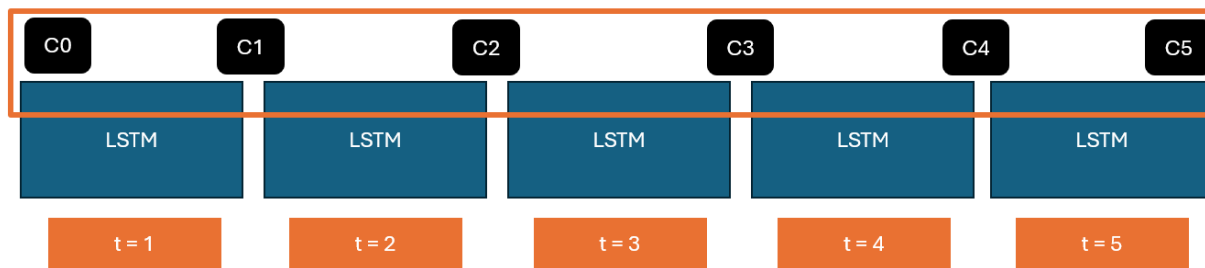
Actualización de la Celda de Estado. Para actualizar la celda de estado se multiplica el valor anterior de la celda de estado por el vector f_t ($f_t \times c_{t-1}$) para eliminar la información que se considera irrelevante, después se crea un vector con los valores candidatos a formar parte de la nueva memoria ($\check{c}_t = \tanh[W_c [a_{t-1}, x_t] + b_c]$), los cuales se multiplican punto a punto con el vector u_t ($u_t \times c_t$) y este resultado se suma a los valores anteriores de la celda de estado ($f_t \times c_{t-1}$) + ($u_t \times \check{c}_t$) (Codificando Bits, 2019).

Nuevo Estado Oculto. Para calcular el nuevo estado oculto se usa la output gate, primero se escala el nuevo valor de la celda de estado en el rango de -1 a 1, con la función tangente hiperbólica, $\tanh [c_t]$. Ahora se usa la output gate para determinar que parte de la celda de estado formará parte del nuevo estado oculto, $o_t = \sigma (W_o [a_{t-1}, x_t] + b_o)$, y finalmente se filtran los valores de la celda de estado con el este vector $o_t \times \tanh[c_t]$ (Codificando Bits, 2019).

Este procedimiento deja de manifiesto la ventaja de una red LSTM de propagar la información almacenada en la celda de estado para cada una de las diferentes celdas que corresponden a un instante de tiempo diferente dentro de una serie, como se muestra en la Figura 5 (Codificando Bits, 2019).

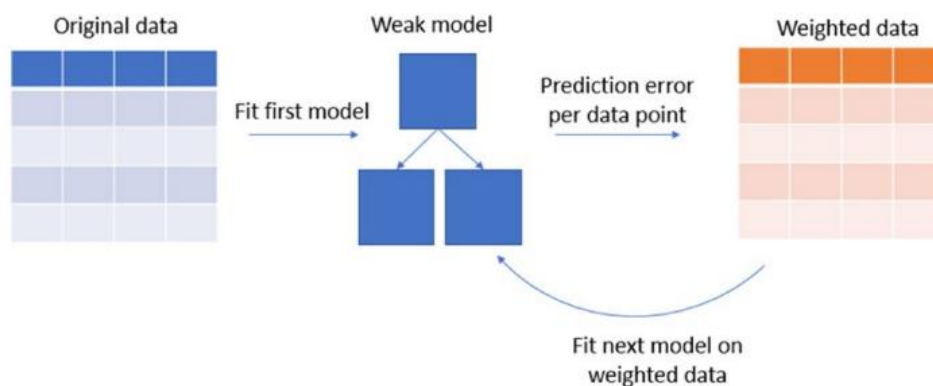
Figura 5

Secuencia Red LSTM



Gradient Boosting

El refuerzo por gradiente es la técnica que logra la combinación de diferentes modelos de árboles de decisión pequeños para hacer predicciones, usando procesos de Boosting y Bagging, los cuales añaden aprendices débiles (weak learners) al modelo de ensemble ponderando los puntos de datos individuales haciendo que cada uno se centre en aprender lo que aún no se entiende, mejorando el conjunto, como se observa en la Figura 6. Con este enfoque de ajuste iterativo de los modelos débiles el ensemble es la combinación de todos los aprendices débiles (Korstanje, 2021, p. 193).

Figura 6*Proceso de Boosting*

Nota. Tomado de (Korstanje, 2021, p. 194).

En síntesis, el objetivo de este tipo de modelos es garantizar que la adición de un nuevo árbol (aprendiz débil) al modelo, explique una cantidad máxima de variación adicional en el modelo de refuerzo de gradiente (Korstanje, 2021, p. 194).

XGBoost

La idea detrás de todos los árboles de decisión es encontrar el mejor punto de división que mejore el modelo y aunque para encontrar este punto basta con iterar varias veces en un bucle con todos los puntos y encontrar el mejor, el costo computacional de este proceso es muy alto, por tanto, XGBoost utiliza una división basada en histogramas. En este caso, el modelo construye histogramas de cada una de las variables y con estos determina cual es la mejor división posible por variable (Korstanje, 2021, p. 195). Cada árbol generará un vector en el espacio de predicción, y en cada nodo tomará decisiones ajustando la etiqueta a medida que se incorpore un nuevo árbol. Este ajuste se realiza restando la suma de las predicciones de los árboles anteriores, ponderada por un factor o tasa de aprendizaje. De esta manera, cada árbol contribuye a corregir los errores cometidos por sus predecesores, mejorando progresivamente.

LightGBM

Este algoritmo usa un método para definir las divisiones denominado GOSS (Gradient Based One Side Sample), calculando los gradientes de cada uno de los puntos de datos y utilizándolos para filtrar los puntos con un gradiente más bajo, excluyendo una proporción significativa de instancias de datos con los gradientes más pequeños, usando el conjunto restante para estimar la ganancia de información. Por otro lado, se emplea el método EFB (Exclusive Feature Bundling), que consiste en agrupar características mutuamente excluyentes con el objetivo de reducir la dimensionalidad del conjunto de datos. Esta optimización disminuye la cantidad de características a procesar, lo que permite acelerar significativamente el proceso de entrenamiento del modelo sin comprometer su precisión (Korstanje, 2021, p. 195; Ke et al., 2017, parr 1).

Prophet

Es un procedimiento automatizado para construir modelos de predicción, que usa un modelo de regresión simple y modular con parámetros predeterminados, permitiendo a los analistas seleccionar los componentes que le sean relevantes de acuerdo con su problema y realizar fácilmente los ajustes necesarios. Y como segundo componente tiene un sistema de medición y seguimiento de la precisión de las predicciones. Por tanto, al ser un modelo real no se tiene una teoría matemática que sirva como aporte (Korstanje, 2021, p. 253; Taylor y Letham, 2017, p. 5).

Estado del Arte

Dentro de los ejemplos de aplicación de técnicas de modelado de datos para la previsión de series de tiempo se encuentran casos como los de (Groene y Zakharov, 2023) en el que usan datos de más de 5 años de ventas por hora de una franquicia de comida rápida en Alemania para

entrenar un algoritmo predictivo, que tiene en cuenta eventos como actividades promocionales, condiciones meteorológicas, fiestas y eventos regionales e indicadores macroeconómicos, con resultados que les permiten mejorar las operaciones optimizando la labor de los gestores de ventas.

En este mismo sentido, la previsión de la demanda de medicamentos es abordada por (Fourkiotis y Tsadiras, 2024), explorando series temporales univariadas, combinando herramientas convencionales como ARIMA y redes neuronales LSTM para refinar la previsión de las ventas, encontrando la eficacia de los modelos de aprendizaje automático en específico el modelo XGBoost, que superó los modelos tradicionales para categorías definidas de medicamentos, con un buen rendimiento de otros modelos como Seasonal Naive y Single Exponential Smoothing.

Otro caso de uso de análisis de series de tiempo se dio para pronosticar la ocupación hotelera empleando datos de ocupación diaria de varias propiedades en entornos urbanos de Estados Unidos con el fin de probar cuatro modelos de pronóstico a corto plazo, lo que dio lugar a concluir que los modelos de Single Exponential Smoothing y XGBoost fueron los más precisos (Ampountolas y Legg, 2023).

Otro caso de estudio fue la investigación de (Shabbir et al., 2023) para el pronóstico de carga eléctrica a corto plazo usando el algoritmo XGBoost con base en la selección de características considerando tendencias en los patrones en los conjuntos de datos históricos, concluyendo que el modelo es capaz de capturar tanto las no lineales y los patrones complejos que se presentan en los datos de carga residencial, superando las técnicas de pronóstico tradicional, logrando mejoras significativas para el pronóstico de carga eléctrica con 24 horas de anticipación.

Ya dentro del análisis de series de tiempo de variables macroeconómicas se encuentra que (Ghosh y Ranjan, 2023) presentan diversas metodologías para el pronóstico del PIB (teniendo en cuenta el retraso con el que se publica este indicador en economías emergentes) usando análisis de series temporales tradicionales como técnicas de aprendizaje automático, incorporando datos de mercados financieros, un índice de incertidumbre económica e indicadores macroeconómicos tradicionales, obteniendo como resultado un mejor rendimiento en la previsión de corto plazo con el uso de un híbrido de ML y métodos convencionales.

En el mismo contexto económico (Mahfooz y Phillips, 2024) precisaron como objetivo pronosticar los precios de Bitcoin durante un mes incorporando variables exógenas (tasa de interés y probabilidad de recesión) y utilizando dos modelos de pronóstico de series de tiempo LSTM y Facebook Prophet entrenaron los modelos con los datos históricos encontrando que LSTM supera a Facebook Prophet en términos de precisión de predicción ya que tiene la capacidad de aprender patrones y relaciones intrincadas debido a su arquitectura de neuronas, destacando que la incorporación de variables exógenas mejoraron significativamente la capacidad predictiva del modelo.

Dentro de la industria fiduciaria se observó la investigación realizada por (Momparler et al., 2024) quienes para determinar la relevancia de los criterios ASG (Principios ambientales, sociales y de gobernanza) en los fondos de acciones registrados en Estados Unidos tomaron los datos de los rendimientos anualizados de los fondos durante un periodo de cinco años (2017-2021), fusionando datos de bases de datos de fondos mutuos con diversas variables entre las cuales se encontraba la calificación en términos de ASG. A partir un enfoque de aprendizaje automático con el algoritmo Extreme Gradient, encontraron la correlación positiva entre las

calificaciones ASG con el desempeño de los fondos, observando que esta variable estaba dentro de las 5 variables principales de explicación del rendimiento.

Dentro del contexto de las administradoras de fondos y fideicomisos de Ecuador (Salgado Gaspar, 2024) desarrolla un modelo estadístico de series temporales para la predicción de los saldos diarios del primer trimestre del año siguiente al cierre del ejercicio, mediante el uso de herramientas estadísticas, el modelo Holt-Winters y la base de datos de saldos diarios de un fondo de inversión de corto plazo concluyendo que el modelo utilizado supera los modelos autorregresivos, sin embargo, recomienda considerar modelos híbridos de modelos autorregresivos con métodos de ML, como redes neuronales, para capturar patrones no lineales en los datos; así como incorporar variables exógenas para mejorar las predicciones incorporando información del entorno macroeconómico.

En el contexto colombiano (Vergara Cardona, 2021) propone un modelo de recomendación de FIC para que una persona de manera autónoma por medio de un cuestionario de perfilamiento de riesgo conecte con el universo de fondos disponibles. Para esto usó algoritmos de clustering que derivaban en la recomendación de una serie de fondos según el perfilamiento de riesgo del usuario.

Metodología

Método

La metodología usada durante el desarrollo de este trabajo se enmarca en lo descrito por (Martínez-Plumed et al., 2021) que menciona a CRISP- DM (Cross-Industry Standard Process for Data Mining) como el estándar para el desarrollo de proyectos de extracción de datos y el descubrimiento de conocimientos más utilizado. Sin embargo, enfatiza en que luego de 20 años desde la introducción de CRISP-DM la diversidad de los datos ha aumentado, así como “la variedad de técnicas de aprendizaje automático, la gestión de datos, la visualización, la inferencia causal y otras áreas” (Martínez-Plumed et al., 2021, p. 1), además del uso del término ciencia de datos como “la extracción de valor de los datos utilizando métodos científicos, como las pruebas de hipótesis estadísticas o el aprendizaje automático” (Martínez-Plumed et al., 2021, p. 4).

Por tal motivo se introduce la metodología cuyos lineamientos siguió este estudio, trayectoria de ciencia de datos (DST por sus siglas en inglés), la cual es compatible con CRISP-DM sin descartarla por completo, sino que la flexibiliza los procesos adaptándola a la realidad actual.

Es por esto por lo que el mapa DST abarca las actividades clásicas de CRISP-DM, sin un orden predeterminado, sino que el orden es establecido por el “líder del proyecto en función de la información disponible, incluidos los resultados de las actividades anteriores” (Martínez-Plumed et al., 2021, p. 5).

Es así como se tienen actividades exploratorias:

1. Exploración de objetivos: encontrar objetivos comerciales que puedan alcanzarse basándose en datos
2. Exploración de fuentes de datos: descubrir fuentes de datos nuevas y valiosas

3. Exploración del valor de los datos: descubrir qué valor se podría extraer de los datos
4. Exploración de resultados: relacionar los resultados de la ciencia de datos con los objetivos del negocio
5. Exploración narrativa: extraer historias valiosas (por ejemplo, visuales o textuales) de los datos
6. Exploración de producto: encontrar formas de convertir el valor extraído de los datos en un servicio o aplicación que ofrezca algo nuevo y valioso a los usuarios y clientes (Martínez-Plumed et al., 2021, p. 7).

Y actividades dirigidas a objetivos CRISP-DM: Entendimiento del negocio, Entendimiento de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue. En la Figura 7 se detalla la secuencia base según la metodología DST.

Cabe aclarar que la trayectoria establecida en un mapa DST no necesariamente sigue un orden predeterminado, ni debe abarcar todas las actividades, además sigue estas consideraciones:

- Un gráfico DST es un gráfico dirigido que solo incluye actividades (una vez) y conexiones (transiciones) entre ellas (como flechas sólidas dirigidas);
- Todas las flechas están numeradas del 0 a N, mostrando la secuencia de transiciones entre actividades. En consecuencia, no podemos tener bucles ilimitados;
- Usamos tres tipos diferentes de cuadros para actividades (círculos para actividades de exploración, cuadrados redondeados para actividades CRISP-DM y cilindros para actividades de gestión de datos);
- Si dos o más flechas tienen el mismo número, significa que ocurren en paralelo (o su orden secuencial no está comprobado o no es importante);

- Una trayectoria puede pasar por la misma actividad más de una vez. Si la trayectoria se mueve de A a B más de una vez, lo anotaremos como una sola flecha con una sola etiqueta, mostrando tantos números de transición como sean necesarios, separados por comas;
- Cada trayectoria tiene una transición de entrada (con número 0 y que no parte de ninguna actividad) y una transición de salida (con número N y no termina en ninguna actividad) (Martínez-Plumed et al., 2021, p. 6).

Procedimiento

Siguiendo la metodología descrita anteriormente y representada en la Figura 7, se establecieron las siguientes fases para la realización de este estudio.

1. Entendimiento del negocio, se llevó a cabo sesiones con expertos en gestión de activos, inversiones y portafolios dentro del sector fiduciario, con el objetivo de identificar las principales variables que influyen en el comportamiento de los activos administrados y las consideraciones para su modelado
2. Definición de objetivos, se utilizó la técnica de árbol de objetivos para establecer el objetivo general y los objetivos específicos del estudio. La estrategia se centra en el desarrollo de modelos de predicción y la selección del mejor modelo con base en métricas de evaluación como RMSE y MAE
3. Adquisición de datos, se recopilaron series históricas tanto de la variable objetivo (activos administrados en los Fondos de Inversión Colectiva) como de las variables exógenas recomendadas por los expertos. Estas series fueron obtenidas de repositorios públicos, asegurando la disponibilidad y trazabilidad de la información
4. Selección de categorías de fondos, se tomaron como base aquellos fondos con los montos más representativos en términos de adiciones y retiros. Se seleccionaron las tres

categorías con mayor peso porcentual, las cuales fueron utilizadas en el proceso de modelado de datos

5. Preprocesamiento de datos, comenzó con la agrupación de la variable objetivo por fecha y la combinación de todas las series temporales mediante esta referencia. Además, se realizó un remuestreo de los datos para garantizar una granularidad diaria uniforme, facilitando su análisis posterior

6. Selección de características, se analizaron las fichas metodológicas y la documentación de las fuentes de datos para descartar variables redundantes o derivadas de combinaciones aritméticas de otras. Este proceso se realizó con base en el conocimiento del negocio, asegurando que solo se incluirán aquellas variables con impacto directo en la variable objetivo

7. Análisis exploratorio de datos; empleó visualizaciones y técnicas estadísticas para identificar patrones estacionales, períodos de comportamiento similar y relaciones entre variables. Esto permitió una comprensión más profunda de la dinámica de los datos y preparar insumos para el modelado

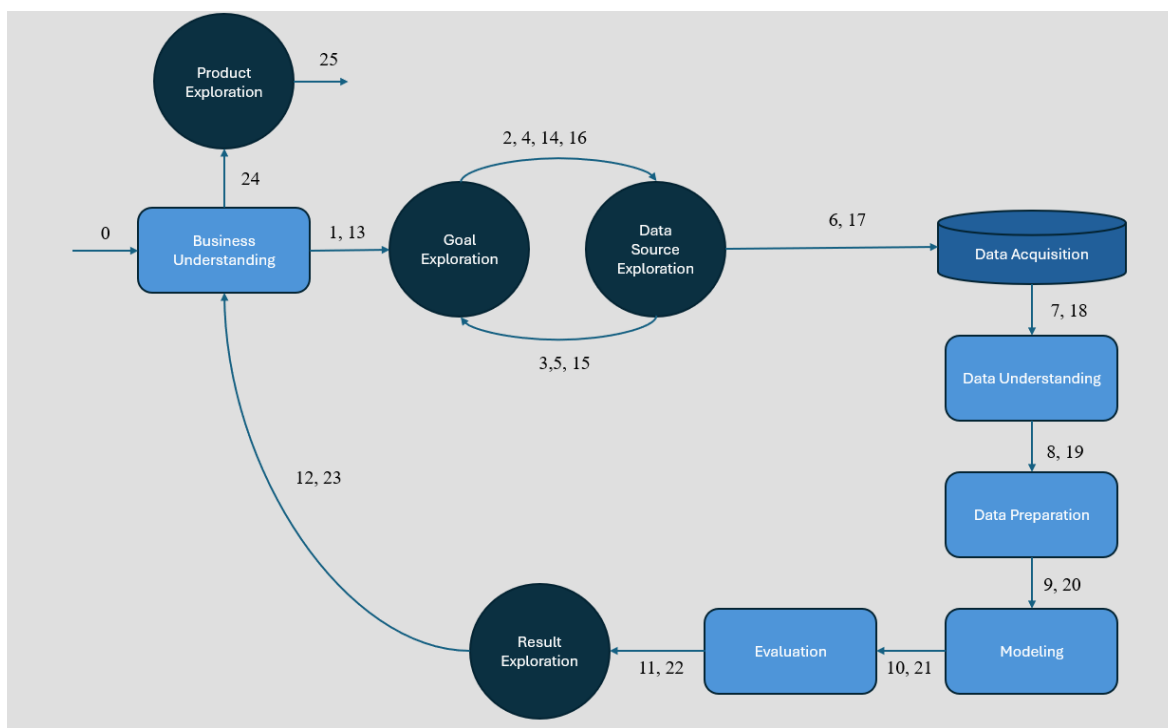
8. Correlación estadística de variables, se aplicaron pruebas de hipótesis para determinar la distribución de las variables y calcular el coeficiente de correlación. Se priorizaron aquellas variables con relaciones moderadas a fuertes con el objetivo de mejorar la capacidad predictiva de los modelos

9. Modelado predictivo, se desarrollaron modelos tanto univariados como multivariados, aplicando diversas técnicas de predicción econométricas y de aprendizaje automático, incluyendo métodos basados en boosting y aprendizaje profundo. Seleccionando los mejores modelos de cada técnica según sus valores de RMSE y MAE

10. Evaluación de Resultados, se compararon los modelos con mejor desempeño de cada técnica predictiva, resaltando aquellos que captaban con mayor precisión las relaciones entre variables y el impacto del tiempo en la variable objetivo. Además, se contrastó el rendimiento del mejor modelo con un modelo simple proporcionando un punto de referencia para evaluar la efectividad de las técnicas avanzadas utilizadas en el estudio

Figura 7

Flujograma Metodología DST



Desarrollo

En este apartado se describen las fases fundamentales para el desarrollo del estudio, iniciando con el entendimiento del contexto mediante las etapas de Business Understanding y Goal Exploration. En esta sección se analizan los factores y variables clave que inciden en el comportamiento de los Fondos de Inversión Colectiva, con base en la perspectiva de expertos de distintas áreas, incluyendo inversión, técnica, financiera y analítica dentro del sector fiduciario. Posteriormente, se abordan las fases de Data Source Exploration, Data Acquisition, Data Understanding y Data Preparation, donde se detallan las fuentes de datos públicos utilizadas para la extracción de las variables identificadas previamente, su caracterización y el análisis de evaluación estadística para determinar su relación con el AUM de los Fondos de Inversión Colectiva. Finalmente, se describe el proceso de almacenamiento y preprocesamiento de los datos, garantizando la calidad y adecuación del conjunto de datos utilizados en la etapa de modelado.

Entendimiento del Negocio

Durante esta fase, se homologó la etapa de definición de requerimientos descrita por (Groene y Zakharov, 2024) mediante la realización de entrevistas con el equipo técnico y operativo de una fiduciaria y del gremio fiduciario. El propósito fue recopilar información relevante sobre el enfoque predictivo actual, los ejercicios previamente realizados en este ámbito y las consideraciones clave para futuras iniciativas de previsión. Para ello, se llevaron a cabo dos reuniones de entendimiento del contexto de negocio. La primera reunión contó con la participación de un profesional del gremio fiduciario, con más de 12 años de experiencia en la gestión de activos y portafolios, análisis financiero, mercado de capitales, modelación financiera y evaluación de proyectos. En la segunda reunión, se entrevistó a dos profesionales de una

fiduciaria con sede en Bogotá: el director de Inversiones, con más de 20 años de trayectoria en el sector financiero, desempeñándose en roles como gerente de producto, miembro del comité de inversiones y director de portafolios; así como un administrador de portafolio en el área de inversiones, con más de 3 años de experiencia en análisis cuantitativo de portafolios de inversión.

En ambas sesiones, se planteó la siguiente pregunta: Desde su experiencia, ¿cuáles consideran que son las variables con mayor incidencia en el comportamiento de los activos administrados por los Fondos de Inversión Colectiva? A partir de este diálogo, se logró identificar una serie de características listadas en la Tabla 1 y factores clave, que influyen en el comportamiento de estos activos, proporcionando una base sólida para la selección de variables.

Tabla 1

Posibles Variables Predictoras

Variable
Agregado monetario M1
Agregado monetario M2
Agregado monetario M3
Tasa de interés de Política Monetaria (TPM)
Índice de Precios al Consumidor (IPC)
Índice de deuda pública
Tasas de cuentas de ahorro
Tasas de captación de Certificados de depósito a término (CDT)
Liquidez de empresas como emisores de valores
Calificación de Colombia en riesgo país
Valor de la Tasa Representativa del Mercado (TRM)

Como resultado del diálogo con los expertos, se identificaron varios factores claves que deben ser considerados en el desarrollo del modelo predictivo para los FIC.

1. Selección del tipo de fondo: Es fundamental elegir un tipo de fondo específico, ya que las características de cada categoría son distintas y los factores que inciden en una pueden diferir significativamente de los que afectan a otra

2. Definición de la ventana de tiempo: Elegir un período temporal alineado con los factores económicos predominantes permite analizar variables exógenas dentro de un contexto económico coherente. Comparar periodos con condiciones económicas disímiles podría distorsionar los resultados del análisis

3. Integración de modelos econométricos clásicos: Se recomienda complementar los modelos de aprendizaje automático con modelos de econometría tradicional, que pueden ofrecer explicaciones sobre las dinámicas económicas que influyen en el saldo de los activos administrados por los FIC. Esto es relevante debido a que los modelos de aprendizaje automático, al ser considerados de tipo caja negra, carecen de interpretabilidad sobre los factores que explican sus resultados

4. Consideración de componentes estacionales: Aunque no ha sido comprobado, la experiencia sugiere que la serie de activos administrados por los FIC presenta un componente estacional que debería ser analizado e incorporado en los modelos predictivos

5. Gestión del rezago en variables exógenas: Al incluir variables exógenas, es necesario considerar el desfase en la publicación de estos datos por parte de distintas fuentes. Indicadores como la inflación suelen divulgarse días después de haber generado efectos directos sobre los FIC, lo cual podría afectar la precisión del modelo si no se gestiona adecuadamente. Por tal motivo, incluir variables con rezago dentro del modelo es una buena opción.

Exploración de Objetivos

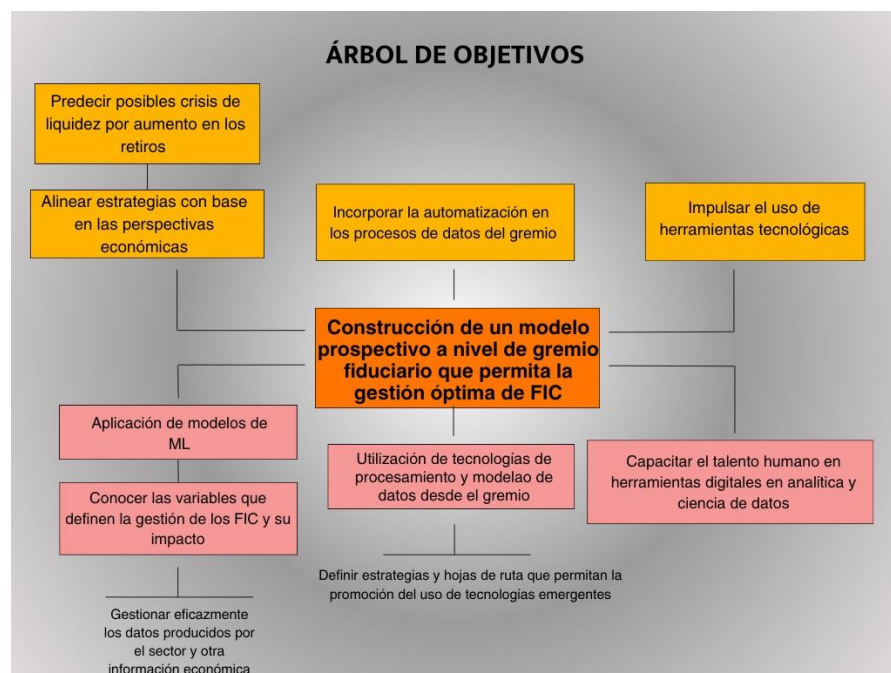
Con base en la información proporcionada por el equipo de expertos, se desarrolló una planificación de las metas a alcanzar en el marco del análisis predictivo de la variable objetivo de activos administrados por los FIC, utilizando la técnica de árbol de objetivos (Barbará, 2021; Santamarina Campos, 2024) útil para en el proceso de planificación en diseños de estrategias de intervención de un problema como lo muestra la Figura 8. El objetivo principal de este análisis fue desarrollar un modelo predictivo que integrara las variables exógenas identificadas con la variable objetivo, con el fin de comprender y anticipar de manera más precisa el comportamiento de los activos administrados por los Fondos de Inversión Colectiva. Para lograrlo, se adoptó un enfoque dual que combinó dos perspectivas complementarias.

En primer lugar, se implementaron modelos econométricos tradicionales, los cuales proporcionan una base sólida para el análisis de relaciones económicas y permiten interpretar el impacto de las variables en la evolución de los activos administrados.

En segundo lugar, se exploró el uso de modelos de Machine Learning, con el objetivo de mejorar la precisión y la capacidad de generalización del modelo predictivo. Estos enfoques permiten captar patrones complejos en los datos, aprovechar un mayor volumen de información y optimizar el desempeño predictivo a través de técnicas avanzadas de entrenamiento y validación.

Figura 8

Diagrama de Árbol de Objetivos para la de Intervención del Problema



Para alcanzar las metas propuestas, se estableció que era necesario definir la categoría de FIC que tuviera la mayor relevancia dentro del universo de Fondos en Colombia, ya fuera por su proporción de activos administrados o por la cantidad absoluta de Fondos. Además, era fundamental caracterizar las variables exógenas para determinar si hay un impacto significativo sobre los activos administrados (Mahfooz y Phillips, 2024), priorizando, en primer lugar, la facilidad para obtener datos históricos y, en segundo lugar, la estadística, seleccionando aquellas variables que presenten una mayor relación con la variable objetivo.

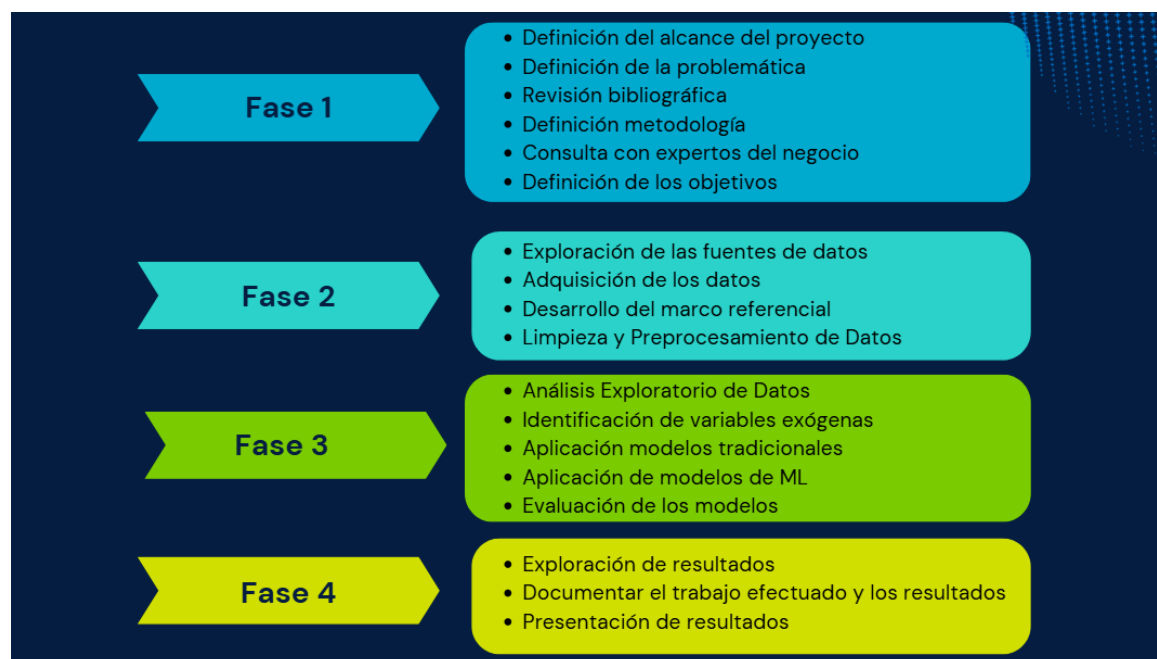
Una vez realizado el análisis de variables y con base en la revisión de la literatura sobre previsión en series de tiempo (Ampountolas y Legg, 2024; Bermejo Iglesias et al., 2009; Groene y Zakharov, 2024; Panchal et al., 2024; Shabbir et al., 2023), se seleccionarían los modelos que hayan demostrado el mejor rendimiento con datos similares a los obtenidos. Estos modelos se

evaluarían mediante métricas como el RMSE (Root Mean Square Error), con el fin de identificar el algoritmo más adecuado para captar la dinámica del mercado de FIC en Colombia, ofreciendo así la mayor precisión en la predicción de los aumentos o descensos de los activos administrados.

Para cumplir con estos objetivos, se siguieron los lineamientos de la metodología representada en la Figura 7, la cual sirvió como guía para la estructuración y ejecución del proyecto. Este marco conceptual permitió planificar de manera detallada las actividades específicas necesarias, como se puede ver en la Figura 9, para el desarrollo del modelo predictivo asegurando un enfoque organizado y coherente.

Figura 9

Fases Ejecución Estudio



Exploración Fuentes de Datos y Extracción de Datos

En esta fase, se realizó una exploración exhaustiva de las principales fuentes públicas de información económica, utilizando como insumo principal el inventario de variables

identificadas durante la fase de Entendimiento del Negocio (Ver Tabla 1), dentro de la exploración se consultaron las fuentes descritas a continuación:

- **Superintendencia Financiera de Colombia:** es el organismo técnico dependiente del Ministerio de Hacienda y Crédito Público que tiene como principales funciones la supervisión de las instituciones financieras en Colombia, promover la estabilidad del sistema financiero, así como proteger a los inversionistas, ahorradores y asegurados. Dentro de su página web periódicamente publica la información estadística del sistema, que consolida a partir de la transmisión que realizan las entidades vigiladas, de allí se extrajo el histórico de AUM FIC del sector fiduciario (Superintendencia Financiera de Colombia, 2024b);

- **Banco de la República de Colombia:** Es el órgano autónomo del Estado colombiano con la función de definir la política monetaria, la política de intervención cambiaria, las políticas crediticias, administrar las reservas internacionales, otorgar liquidez transitoria al sistema financiero, emitir la moneda legal colombiana, así como contribuir al buen funcionamiento de los sistemas de pago y generar conocimiento mediante la investigación económica y financiera, además de la producción y difusión de estadísticas. Dentro de su página web se extrajo la serie histórica de tasas de interés de política monetaria, la serie histórica de agregados monetarios crediticios y la serie histórica de tasas de captación de CDT (Banco de la República de Colombia, 2024a, 2024b, 2024c);

- **Departamento Administrativo Nacional de Estadística (DANE):** Esta entidad perteneciente a la rama ejecutiva del poder público es la encargada de recopilar, procesar y difundir información estadística oficial en Colombia, garantizando que cumpla con los estándares internacionales. Mensualmente publican la cifra del Índice de Precios al Consumidor

(IPC), por lo tanto de su página web se extrajo esta serie histórica (Departamento Nacional de Estadística, 2024);

- Bolsa de Valores de Colombia: es la entidad privada que se encarga de operar la infraestructura del mercado de capitales en Colombia, desde la negociación de los títulos hasta su custodia en el servicio de depósito. Diariamente pone a disposición del público diferentes cifras estadísticas del mercado de capitales, entre ellas las de divisas e índices representativos de deuda, por tal motivo de su página web se extrajo la serie histórica del valor del dólar y la serie histórica del valor del índice del desempeño de los títulos de deuda pública emitidos por el gobierno colombiano (COLTES) (Bolsa de Valores de Colombia, 2024a, 2024b);

- Sistema de Categorización de Fondos de Inversión Colectiva (SIFIC): este sistema fue el resultado del esfuerzo conjunto de la Asociación de Fiduciarias, la Asociación de Comisionista de Bolsa y LVA Índices para establecer grupos de FIC, haciéndolos comparables entre sí para que el inversionista conozca los fondos y pueda evaluarlos. Periódicamente a través de la información que las entidades le envían publica las rentabilidades y estadísticas de los FIC, por tanto de su página web extrajo la serie histórica de rentabilidades de los FIC Categorizados (LVA Colombia, 2024).

Estas cinco fuentes de información, representadas en la Figura 10, constituyen la base fundamental para la recopilación de datos relevantes, alineados con los factores clave identificados por el equipo de expertos durante la etapa de Entendimiento del Negocio. La integración de estos datos en los modelos predictivos permite realizar un análisis integral del comportamiento de los activos administrados por los FIC.

Las variables seleccionadas sintetizan y describen el comportamiento del portafolio de inversiones de los FIC, los cuales destinan aproximadamente el 76% de sus recursos a

instrumentos representativos de deuda de emisores nacionales de renta fija (Superintendencia Financiera de Colombia, 2024a), como Títulos de Tesorería, otros títulos de deuda pública, Títulos de Devolución de Impuestos (TIDIS), Títulos de Desarrollo Agropecuario (TDA), Títulos de Solidaridad (TDS), Bonos Pensionales, Bonos Ordinarios, Bonos Subordinados, Títulos Corporativos de Participación, Papeles Comerciales y Titularizaciones (Bolsa de Valores de Colombia, 2024c). Por otro lado, el 9.2% de las inversiones se dirige a instrumentos representativos de deuda de emisores extranjeros, el 6.8% a instrumentos de patrimonio de emisores nacionales y el 6.1% a Títulos de Tesorería (Superintendencia Financiera de Colombia, 2024a).

La relevancia de estas variables radica en su capacidad para reflejar el comportamiento de los FIC en el mercado financiero colombiano, permitiendo a los modelos predictivos identificar tendencias, patrones y relaciones clave que pueden influir en la gestión de los activos administrados por los FIC.

Figura 10

Resumen Fuentes de Datos



Entendimiento de los Datos

El proceso de comprensión del conjunto de datos se centra en analizar su estructura, identificando las características que lo componen, su formato y la definición de cada variable incluida. Además, se establecieron las fechas de inicio y fin de las series, lo que permitió comprender el alcance temporal de la información disponible. Un aspecto clave en esta fase fue la evaluación de la granularidad de los datos, es decir, la unidad temporal con la que fueron registradas, lo cual resultó fundamental para definir la frecuencia adecuada del pronóstico (Groene y Zakharov, 2024).

Este análisis permitió determinar si los datos requerían algún tipo de transformación o consolidación para garantizar su coherencia temporal con el objetivo del estudio. Asimismo, facilitó la identificación de posibles inconsistencias, valores faltantes o irregularidades en las series temporales, lo que sirvió como base para la posterior etapa de preparación de datos, asegurando la calidad de la información utilizada en el modelado predictivo.

En la Tabla 2, se detalla la estructura de cada conjunto de datos disponible, tras la caracterización de las series históricas obtenidas de las fuentes previamente identificadas.

Tabla 2

Descripción Conjuntos de Datos

Conjunto de Datos	VARIABLES REPRESENTATIVAS	Tipo de Variables	Granularidad
AUM FIC	Fecha, Subtipo, cod_entidad, nombre_negocio, cod_negocio, AUM FIC, Inversionistas	Datetime, Categórica, Numérica	Diaria
Agregados Monetarios y Crediticios	Fecha, Efectivo, Reserva Bancaria, Total del sector privado, Del sector público, Cuentas de ahorros, Saldo	Datetime, Numérica	Semanal

CDT, M1, M2, M3, Depósitos restringidos			
Índice de Precios al Consumidor	Fecha, Índice, Inflación anual %, Inflación mensual %, Inflación año corrido %	Datetime, Numérica	Mensual
Tasa de Intervención de política monetaria	Fecha, Tasa de intervención de política monetaria (%)	Datetime, Numérica	Mensual
Categorización FIC	Fecha, Fiduciaria, Código_negocio, Categoría, Nombre_participacion	Datetime, Categórica	Diaria
Tasas de Captación CDT	Fecha, Tasas y Montos entre 30 y 365 días	Datetime, Numérica	Semanal
Valor COLTES	Fecha, Valor hoy	Datetime, Numérica	Diaria
Valor dólar	Valor, Vigencia desde, Vigencia hasta	Datetime, Numérica	Diaria

La selección de características se llevó a cabo, usando las fichas metodológicas (Banco de la República de Colombia, 2020, 2021; Bolsa de Valores de Colombia, 2020; Departamento Nacional de Estadística, 2019) o la descripción del conjunto de datos disponible en cada una de las fuentes públicas consultadas para definir los siguientes criterios de selección: eliminar del análisis las variables que fueran una combinación de otras, por ejemplo, el Agregado Monetario M1 es la suma de Efectivo y los depósitos en cuentas corrientes, por tal razón se dejaron las variables individuales dejando por fuera M1; el segundo criterio fue el aporte directo a la variable objetivo, como es el caso de Bonos en poder del banco de la República, ya que no hacen parte de los agregados monetarios (Banco de la República de Colombia, 2020).

Como resultado de este ejercicio de selección de características, fundamentado en el entendimiento del negocio, se identificaron 28 variables con potencial relevancia como

explicativas del comportamiento de los activos administrados en los FIC. De este modo, se excluyeron 46 variables. Las variables seleccionadas se listan en la Tabla 3.

Tabla 3

Variables Seleccionadas Después del Filtrado de Variables

Variable	Tipo de Variable
Fecha	Datetime
AUM FIC	Numérica
Reserva Bancaria	Numérica
Del sector privado	Numérica
Del sector público	Numérica
Cuentas de ahorro	Numérica
CDT	Numérica
Otros depósitos a la vista	Numérica
Inflación anual	Numérica
Tasa de intervención de política monetaria	Numérica
Tasa a 30 días	Numérica
Tasa Entre 31 y 44 días	Numérica
Tasa a 45 días	Numérica
Tasa Entre 46 y 69 días	Numérica
Tasa a 60 días	Numérica
Tasa Entre 61 y 89 días	Numérica
Tasa A 90 días	Numérica
Tasa Entre 91 y 119 días	Numérica
Tasa A 120 días	Numérica
Tasa Entre 121 y 179 días	Numérica
Tasa A 180 días	Numérica
Tasa Entre 181 y 359 días	Numérica
Tasa A 360 días	Numérica
Tasa Superiores a 360 días	Numérica

Tasa Captaciones a través de CDT por red de oficinas	Numérica
Tasa Captaciones a través de CDT por tesorería	Numérica
COLTES	Numérica
Valor dólar	Numérica

Preparación de Datos

La etapa de preparación o preprocesamiento de datos fue fundamental para garantizar la calidad y utilidad de la información utilizada en los modelos econométricos y de aprendizaje automático. Su propósito principal fue optimizar los datos, asegurando que presentarán la estructura adecuada y estarán libres de errores que podrían afectar el desempeño de los modelos. Para este proceso, se aplicó la metodología propuesta por (Mertz, 2021) la cual integra diversas técnicas orientadas a mejorar la integridad de los datos. Se llevó a cabo la homologación de formatos, estandarizando tipos de datos, fechas y unidades de medida para asegurar coherencia en todas las variables.

Para llevar a cabo esta tarea, se identificó la característica temporal de cada conjunto de datos, transformándola en un objeto de tipo `datetime` (Pandas, 2024b). Este proceso permitió establecer los puntos de tiempo sobre los cuales se analizarían las observaciones cuantitativas de las características definidas en la Tabla 3, logrando así estructurar cada conjunto en forma de serie temporal (D. A. Pal y Prakash, 2017).

Posteriormente, a cada conjunto de datos se le aplicó un método de remuestreo (Pandas, 2024a) basado en la característica temporal, ajustando la frecuencia a una escala diaria. Este ajuste garantizó la compatibilidad de todas las series temporales con la frecuencia del conjunto principal, el AUM FIC. De esta manera, se aseguró que todas las series pudieran fusionar o integrar correctamente para su análisis conjunto.

El tratamiento de valores fue otra acción clave dentro del proceso, comenzando con la implementación de estrategias de identificación de anomalías en datos tabulares, basadas en los diez patrones de anomalías más comunes descritos por (Sukhobok et al., 2017) Entre estas anomalías se chequeó la presencia de valores incorrectos, valores inconsistentes (ver figura 11), encabezados de columnas con información de atributos, encabezados incorrectos, columnas y filas no relacionadas con el modelo de datos, múltiples valores almacenados en una sola columna, valores únicos a lo largo de múltiples columnas y valores duplicados. Este último fue el factor determinante para la eliminación de información no útil. Además, se realizó la identificación de valores nulos, los cuales se imputaron mediante el método de relleno hacia adelante (forward fill, FFill), garantizando la continuidad y calidad de las series temporales.

Para evitar sesgos relacionados con las escalas de medición, que pueden impactar el desempeño de algunos modelos predictivos utilizados en este estudio, se implementó la técnica de normalización Min-Max (Zheng y Casari, 2018) Esta decisión se tomó considerando que, como se detalla en la sección de caracterización de variables, las distribuciones no presentan un comportamiento normal y no se evidenció una fuerte presencia de valores atípicos. Al optar por esta técnica, se garantizó la conservación de la proporción entre las variables, ajustándolas a rangos compatibles entre sí (Mertz, 2021).

Figura 11

Identificación Gráfica de Valores Inconsistentes



Nota. Se identificó un valor inconsistente en el registro de marzo de 2022, el cual se desvía significativamente de la tendencia observada en la serie temporal. Esta anomalía se debe a la duplicidad del mismo valor

Se realizó una validación integral de la coherencia y consistencia de los datos procesados, garantizando que el conjunto final cumpliera con los requisitos técnicos para alimentar los modelos predictivos. Este enfoque permitió obtener datos limpios, estructurados y alineados con los objetivos del análisis, contribuyendo significativamente a la precisión y confiabilidad de los resultados.

Una actividad relevante dentro de esta etapa fue la definición de la ventana temporal para el análisis, siguiendo la recomendación del equipo de expertos. Se seleccionó el período comprendido entre enero de 2018 y agosto de 2024, considerando que este intervalo permite capturar distintas dinámicas del mercado, influenciadas tanto por tendencias financieras como por cambios en políticas económicas y tributarias (Herrera y Angarita, 2023; Grupo Proindustria, 2024).

Este rango temporal se divide en tres periodos, como los observados en la Figura 12, cada uno con comportamientos diferenciados: 2018-2020, se observa un crecimiento sostenido pero moderado, reflejando estabilidad en los mercados y condiciones económicas relativamente favorables; 2020-2022, este período destaca por la abrupta caída del primer trimestre de 2020, atribuida a la crisis generada por la pandemia, seguida de una recuperación gradual hasta 2021 y una nueva contracción hacia 2022 (Herrera y Angarita, 2023); 2022-2024: Se caracteriza por un crecimiento acelerado, impulsado por la recuperación económica global y la implementación de nuevas políticas económicas y tributarias (Grupo Proindustria, 2024).

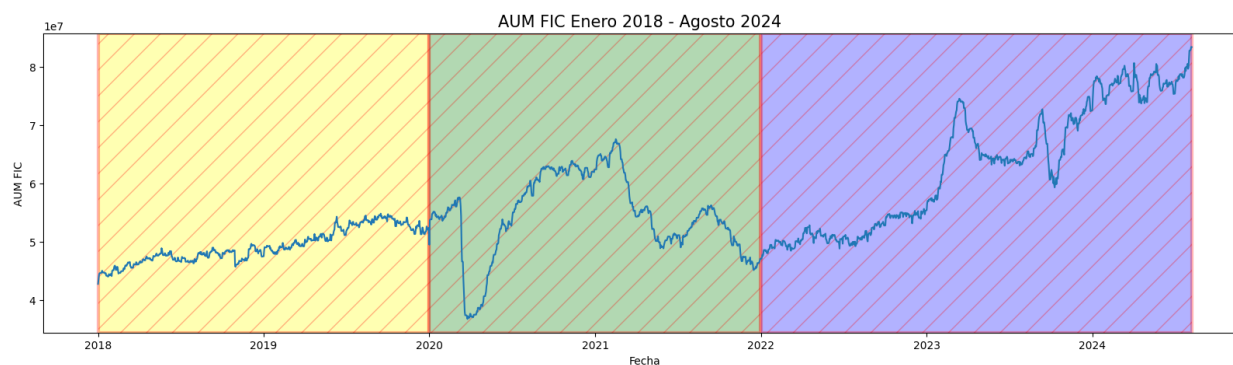
Como describe (Herrera y Angarita, 2023) :

El cuatrienio 2018-2022 presentó una intensa política tributaria, definida por dos etapas. La primera corresponde a la implementación de la visión de política económica del Gobierno nacional, enmarca en impulsar el crecimiento económico por medio de una disminución de los impuestos a las sociedades, la reducción del costo del uso del capital y la eliminación de distorsiones que afectaran la inversión privada. La segunda etapa se enmarca en la pandemia del COVID -19, con medidas para hacer frente a la situación de emergencia económica y social, y – posteriormente – con medidas para corregir las finanzas públicas e impulsar la recuperación económica. (p.18)

En este mismo sentido de etapas, el periodo 2022-2024, muestra una etapa con una economía estancada, creciendo a menos de 2%, con altos costos de financiamiento, bajos niveles de confianza, con un déficit en la balanza comercial, reducción de la inflación, aumento de y posterior recorte sostenido de tasas de interés, un riesgo en la sostenibilidad fiscal y empleo y desempleo persistentes (Grupo Proindustria, 2024).

Figura 12

Serie Histórica AUM FIC Fiduciarias



Nota. Se muestra la serie para el periodo enero 2018 y julio 2024

También se destaca la selección de categorías de fondos, una recomendación clave proporcionada por el equipo de expertos. Esta actividad fue esencial para enfocar el análisis en los segmentos más representativos del mercado, considerando aquellos fondos con mayor volumen de aportes y retiros durante el periodo seleccionado (Datos Abiertos de Colombia, 2024). Se excluyeron los fondos sin categoría.

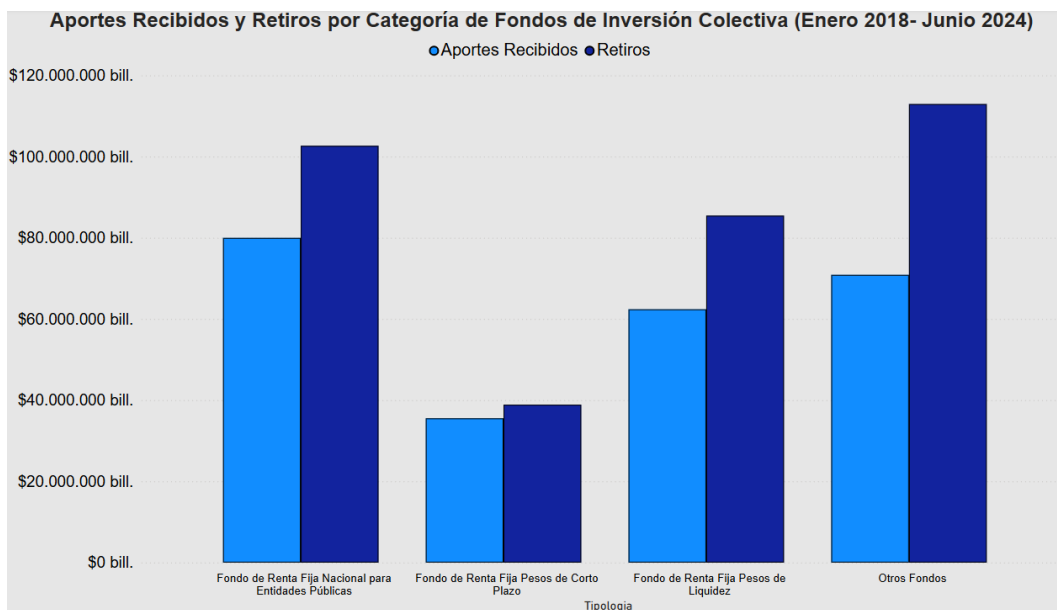
La elección de estas categorías se fundamentó en su peso significativo dentro de todos los FIC administrados por sociedades fiduciarias, dado que concentran una parte considerable de los flujos de capital, convirtiéndose en el foco principal de los modelos predictivos, como lo muestra la Figura 13. Este enfoque busca mejorar la precisión en la proyección de los activos administrados, al centrarlo en los fondos que tienen mayor impacto en la dinámica del mercado.

De acuerdo con la clasificación establecida por (SIFIC, 2021) las categorías seleccionadas para filtrar la información sobre los Activos Bajo Administración (AUM) de los FIC fueron:

- Fondos de Renta Fija Nacional para Entidades Públicas, aquellos FIC con una duración total del valor del fondo (incluyendo disponible) menor o igual a 365 días, con políticas de inversión que cumplan las disposiciones del Decreto 1525 de 2008 y sus modificatorios y que, a solicitud de la Entidad Participante que los administra, sean incluidos en esta categoría;
- Fondos de Renta Fija Pesos de Liquidez, aquellos FIC de Renta Fija con política de inversión conservadora, cuyo objetivo principal sea preservar el capital, con una duración total del valor del fondo menor o igual a 240 días, y al menos un 80% de su portafolio invertido en instrumentos cuya calificación nacional sea superior o igual a la segunda más alta vigente en el largo plazo según la escala utilizada por las sociedades calificadoras en instrumentos con plazo superior a un (1) año; o cuya calificación nacional sea igual a la máxima vigente en el corto plazo de acuerdo con la escala utilizada para este plazo en instrumentos con plazo inferior o igual a un (1) año; o inversiones expuestas a riesgo nación;
- Fondo de Renta Fija Pesos de Corto Plazo, FIC que por sus objetivos de inversión mantengan una duración total del valor del fondo en el rango de los 240 y los 540 días, y al menos un 80% de su portafolio invertido en instrumentos cuya calificación nacional sea superior o igual a la segunda más alta vigente en el largo plazo según la escala utilizada por las sociedades calificadoras en instrumentos con plazo superior a un (1) año; o cuya calificación nacional sea igual a la máxima vigente en el corto plazo de acuerdo con la escala utilizada para este plazo en instrumentos con plazo inferior o igual a un (1) año; o inversiones expuestas a riesgo nación.

Figura 13

Aportes Recibidos y Retiros por Categoría de FIC



Nota. Se muestran los datos para el periodo enero 2018- junio 2024

Figura 14

Actividades de Preprocesamiento de Datos



Caracterizaciones Variables Económicas

Uno de los objetivos fundamentales de este estudio fue identificar las variables con mayor influencia sobre el comportamiento de los activos administrados por las Sociedades Fiduciarias a través de los Fondos de Inversión Colectiva (FIC). Para lograrlo, se emplearon análisis de evaluación estadística con el propósito de determinar qué variables presentaban una relación significativa con la variable objetivo.

Dado este enfoque, se consideran como métodos candidatos el Coeficiente de Correlación de Pearson y el Coeficiente de Correlación de Spearman, ambos adecuados para medir el grado de asociación entre variables cuantitativas, aunque con características metodológicas distintas. La elección entre estos dos métodos dependía directamente de la distribución de los datos: el Coeficiente de Correlación de Pearson es un método paramétrico que mide la relación lineal entre dos variables, siempre y cuando se cumple el supuesto de normalidad univariante. Esto implica que las variables deben ajustarse a una distribución normal para que el análisis sea válido (Lizama et al., 2014); y el Coeficiente de Correlación de Spearman, por su parte, es un método no paramétrico que evalúa la relación monótona entre variables, sin exigir que los datos sigan una distribución específica. Esto lo hace especialmente útil cuando los datos presentan asimetrías, valores atípicos o no cumplen con los supuestos de normalidad. (Lizama et al., 2014)

Por ello, antes de seleccionar el método de comparación más adecuado, se realizó un análisis de la distribución de las variables mediante pruebas de normalidad. Para llevar a cabo este análisis se eligió la prueba de Anderson -Darling (Kvam y Vidakovic, 2007) usada para verificar si una muestra de datos pertenece a una distribución específica, con la ventaja que le otorga más peso a las colas de dicha distribución, haciéndola más sensible a desviaciones de la distribución, con la ventaja que tiene un mejor comportamiento con tamaños de muestra

moderados a grandes (SixSigma, 2024). Para aplicar esta prueba se utilizó el software libre especializado en análisis de datos con capacidades de estadística univariada y multivariada, PAST versión 5.0.1 (Natural History Museum, s.f.). Este programa facilitó la aplicación de la prueba de normalidad de Anderson- Darling, reconocida por su eficacia en la detección de desviaciones de la normalidad (Tapia y Cevallos, 2021). La prueba de Anderson- Darling permitió determinar si las series temporales de las variables analizadas seguían una distribución normal, como se muestra en la Tabla 4. Los resultados de esta prueba fueron determinantes para la elección del coeficiente de compensación más apropiada ya que si las variables mostraban una distribución normal, se optaba por el uso del Coeficiente de Correlación de Pearson, dada su capacidad para medir relaciones lineales bajo este supuesto. En caso contrario, se seleccionaba el Coeficiente de Correlación de Spearman, por ser un método no paramétrico que no requiere normalidad y es más robusto frente a distribuciones asimétricas o presencia de valores atípicos.

Tabla 4

Resultados Prueba de Anderson - Darling

Variable	p valor	Normalidad
AUM FIC	1,81E-150	Los datos no siguen una distribución normal
Efectivo	0,00E+00	Los datos no siguen una distribución normal
Reserva Bancaria	6,12E-51	Los datos no siguen una distribución normal
Del sector privado	2,23E-146	Los datos no siguen una distribución normal
Del sector público	3,49E-34	Los datos no siguen una distribución normal

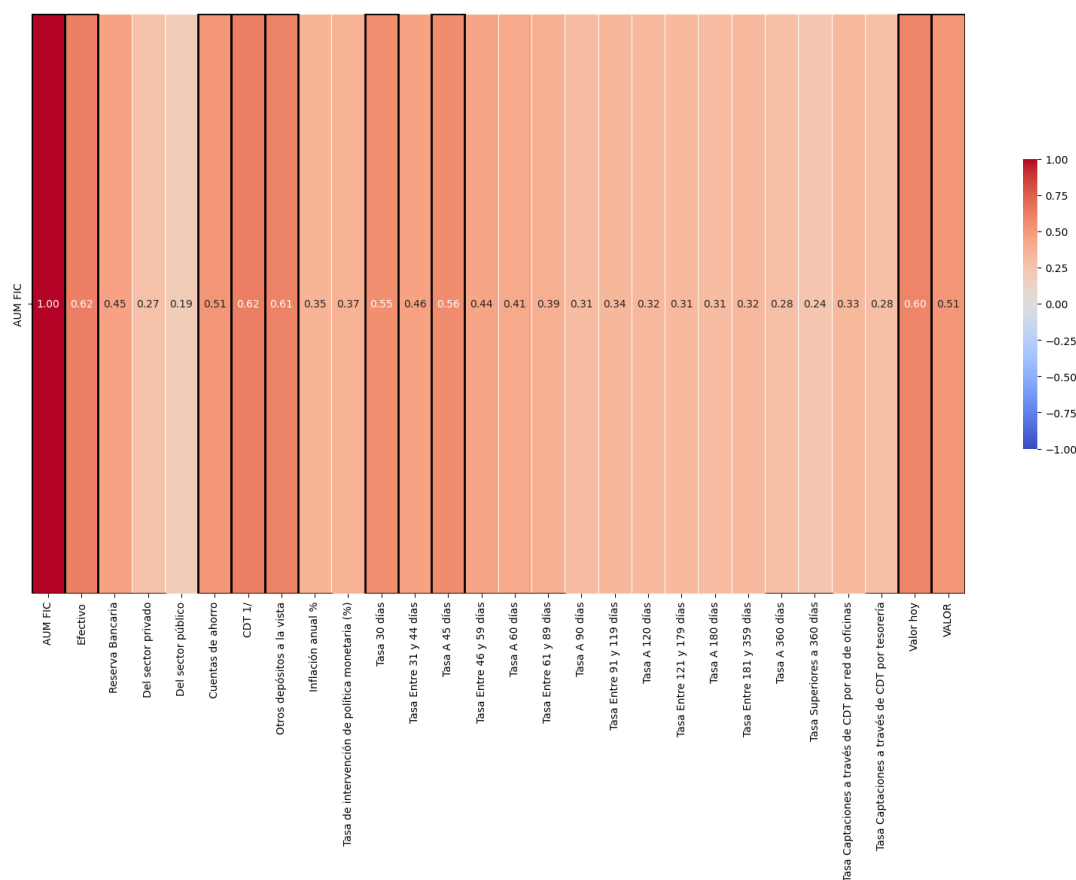
Cuentas de ahorro	0,00E+00	Los datos no siguen una distribución normal
CDT 1/	0,00E+00	Los datos no siguen una distribución normal
Otros depósitos a la vista	1,37E-136	Los datos no siguen una distribución normal
Inflación anual %	0,00E+00	Los datos no siguen una distribución normal
Tasa de intervención de política monetaria (%)	0,00E+00	Los datos no siguen una distribución normal
Tasa 30 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 31 y 44 días	0,00E+00	Los datos no siguen una distribución normal
Tasa A 45 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 46 y 59 días	0,00E+00	Los datos no siguen una distribución normal
Tasa A 60 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 61 y 89 días	0,00E+00	Los datos no siguen una distribución normal
Tasa A 90 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 91 y 119 días	0,00E+00	Los datos no siguen una distribución normal
Tasa A 120 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 121 y 179 días	0,00E+00	Los datos no siguen una distribución normal

Tasa A 180 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Entre 181 y 359 días	0,00E+00	Los datos no siguen una distribución normal
Tasa A 360 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Superiores a 360 días	0,00E+00	Los datos no siguen una distribución normal
Tasa Captaciones a través de CDT por red de oficinas	0,00E+00	Los datos no siguen una distribución normal
Tasa Captaciones a través de CDT por tesorería	0,00E+00	Los datos no siguen una distribución normal
Valor COLTES	2,18E-102	Los datos no siguen una distribución normal
Valor dólar	5,46E-36	Los datos no siguen una distribución normal

Al comprobar que ninguna de las variables sigue una distribución normal debido a que el p-valor se encuentra por debajo de 0.05, rechazando la hipótesis nula de que el conjunto de datos sigue una distribución normal, se optó por el cálculo del coeficiente de correlación de Spearman, cuyos resultados se muestran en la Figura 15.

Figura 15

Matriz de Correlación de Variables



El análisis del coeficiente de correlación revela que varias variables presentan una relación significativa con los activos administrados de los FIC. Específicamente, se identifican variables con una calificación moderada (entre 0.4 y 0.6) y otras con una calificación fuerte (entre 0.6 y 0.8), ordenadas de forma descendente según su grado de asociación: Efectivo, Certificados de Depósito a Término (CDT), Otros Depósitos a la Vista, Valor del índice COLTES, Tasas de CDT a 45 y 30 días, Cuentas de ahorro y Valor del dólar.

El hecho de que todos los coeficientes de calificación sean positivos indica que un incremento en cualquiera de estas variables está asociado con un aumento en los activos administrados por los FIC. Esto sugiere que estos factores podrían desempeñar un papel

relevante en la toma de decisiones de inversión y en la dinámica de crecimiento de los fondos, representando variables clave para el desarrollo de modelos predictivos.

Modelado de Datos

El propósito de esta fase es detallar el proceso empleado en la etapa de modelado de datos, en el cual se entrenaron veintidós modelos predictivos de series de tiempo utilizando distintas técnicas, incluyendo ARIMA, ARIMAX, Suavizado Exponencial, VAR, XGBoost, LightGBM, redes LSTM y Prophet. Cada uno de estos modelos fue ajustado con diversas configuraciones de parámetros, los cuales se presentan en diagramas de flujo para ilustrar el procedimiento de selección. La elección de la mejor configuración para cada modelo se basó en un análisis comparativo de métricas de evaluación, tales como RMSE y MAE, obteniendo 9 modelos (uno por cada técnica de predicción) garantizando así la selección del modelo con el mejor desempeño predictivo.

Con base en los estudios realizados por (Fourkiotis y Tsadiras, 2024), (Ampountolas y Legg, 2023), (Shabbir et al., 2023), (Ghosh y Ranjan, 2023), (Mahfooz y Phillips, 2024) y (Salgado Gaspar, 2024) se identificaron varios modelos que han demostrado un rendimiento superior en tareas de pronóstico, especialmente en contextos que integran variables macroeconómicas. Entre estos modelos destacan ARIMA, VAR, Suavizado Exponencial, Redes LSTM, XGBoost, LightGBM y Prophet, los cuales han sido utilizados por su capacidad para capturar patrones complejos en series de tiempo y adaptarse a distintos escenarios, entre ellos el económico.

Considerando estas referencias y con los datos disponibles, se procedió a entrenar y evaluar diversos ajustes de estos modelos. Para cada uno, se determinó la combinación óptima de

parámetros que mejor se ajustaba al contexto específico de los Fondos de Inversión Colectiva (FIC), teniendo en cuenta la naturaleza de los datos y las dinámicas del mercado.

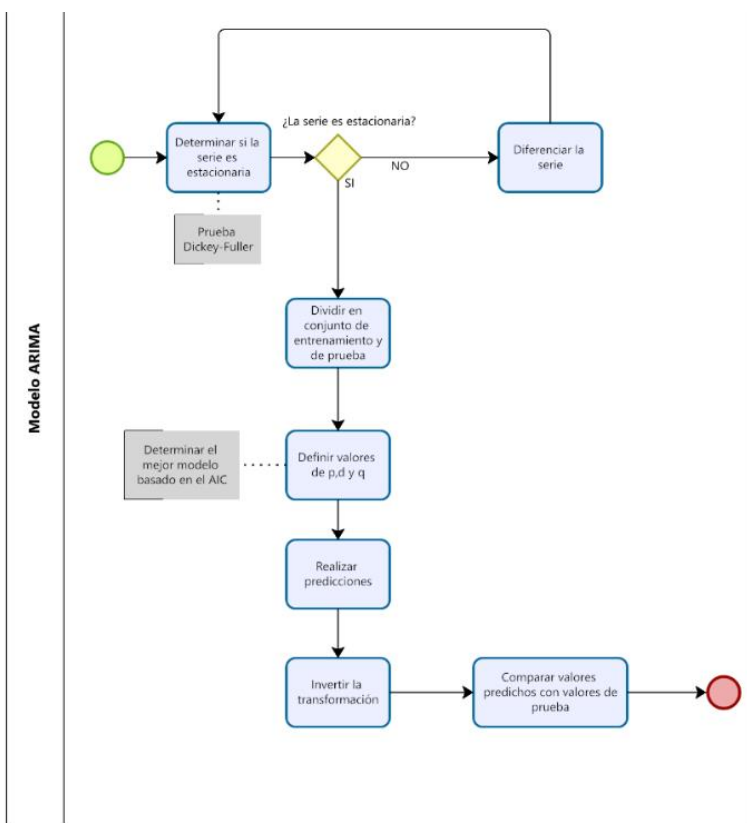
Como resultado de este proceso, se seleccionaron ocho modelos que fueron evaluados de manera rigurosa utilizando métricas de desempeño como la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE). Estas métricas permitieron comparar objetivamente la precisión de cada modelo, con el propósito de identificar aquel que ofrece el mejor ajuste a los datos y las predicciones más precisas.

Modelo ARIMA

Para el modelo ARIMA se siguieron los pasos descritos en la Figura 16.

Figura 16

Diagrama de Flujo Procedimiento Modelo ARIMA



El modelo seleccionado fue el ARIMA (2,4,3), cuyos resultados se presentan en la Tabla 5. Este modelo obtuvo métricas de evaluación de RMSE: 3.92×10^6 y MAE: 3.81×10^6 , calculadas a partir de la comparación entre las predicciones generadas y los valores del conjunto de prueba. La representación gráfica de esta comparación se muestra en la Figura 17, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

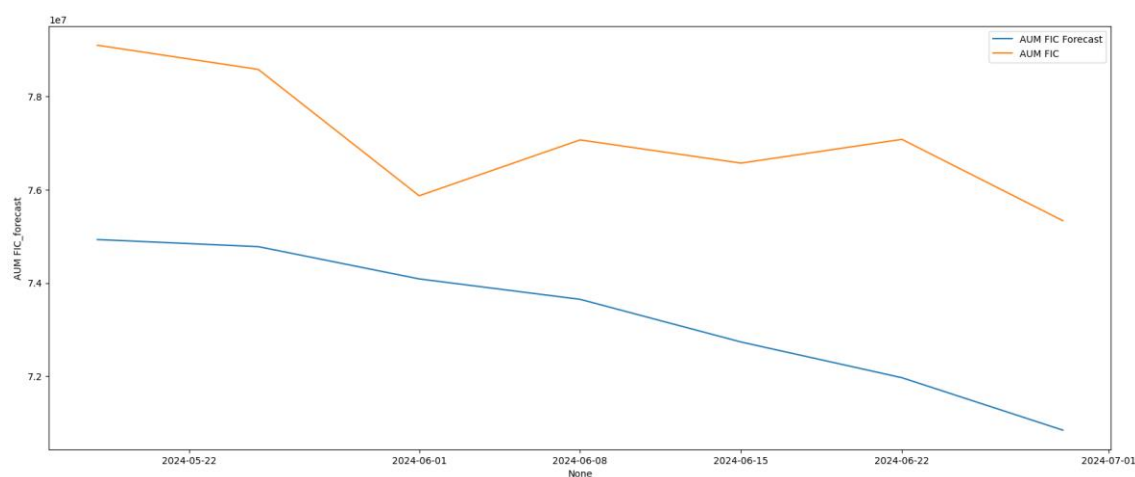
Tabla 5

Resultados Modelo ARIMA

Variable	Coefficiente	Std err	Z	P> z	[0.025	0.975]
ar.L1	-1.5739	0.011	-144.168	0.000	-1.595	-1.553
ar.L2	-0.6574	0.016	-39.982	0.000	-0.690	-0.625
ma.L1	-0.9966	0.339	-2.942	0.003	-1.661	-0.333
ma.L2	-0.9999	0.340	-2.940	0.003	-1.666	-0.333
ma.L3	0.9967	0.015	66.984	0.000	0.968	1.026
Sigma2	2.80e+10	3.63e+11	7.73e+20	0.000	2.81e+10	2.81e+10

Figura 17

Gráfico AUM Dato Real vs Predicción ARIMA

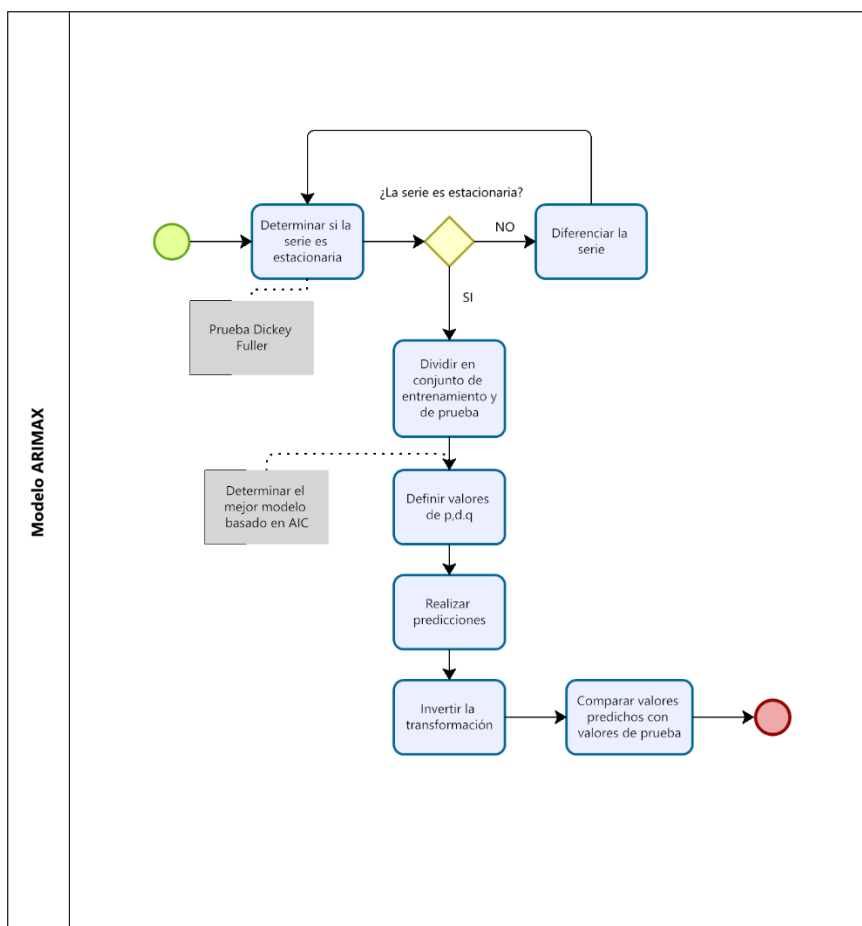


Modelo ARIMAX

Como variable exógena para este modelo se eligió la variable Efectivo, ya que es la variable con el coeficiente de correlación más alto, como se muestra en la Figura 15. Para el modelo ARIMAX se siguieron los pasos descritos en la Figura 18.

Figura 18

Diagrama de Flujo Procedimiento Modelo ARIMAX



El modelo entrenado fue el ARIMAX (2,3,3) cuyos resultados se presentan en la Tabla 6. Este modelo obtuvo métricas de evaluación de RMSE: 2.16×10^6 y MAE: 1.75×10^6 , calculadas

a partir de la comparación entre las predicciones generadas y los valores del conjunto de prueba.

La representación gráfica de esta comparación se muestra en la Figura 19, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

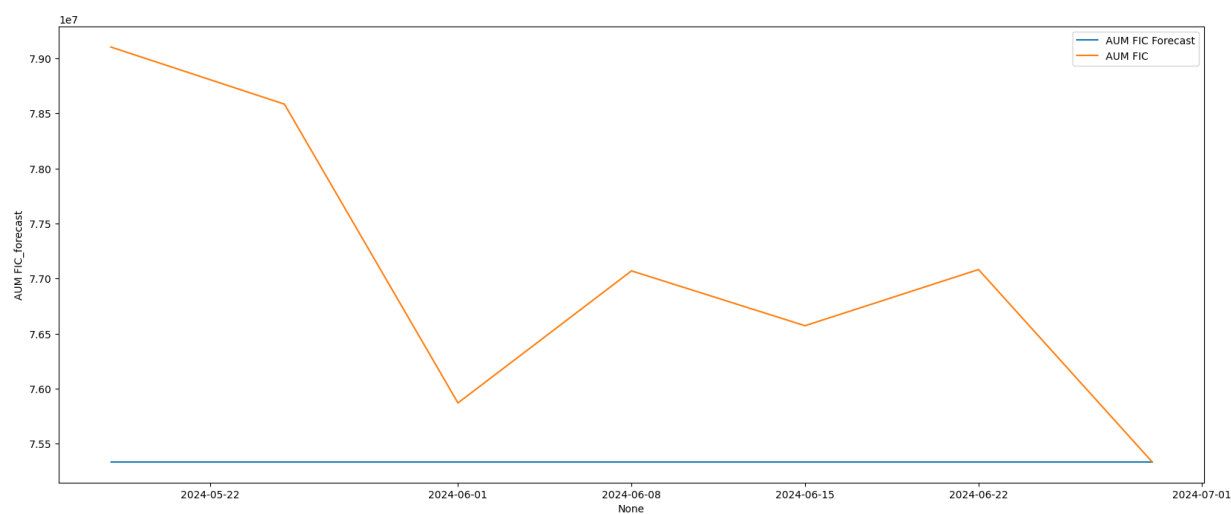
Tabla 6

Resultados Modelo ARIMAX

Variable	Coefficiente	Std err	Z	P> z	[0.025	0.975]
AUM FIC	1.000	2.28e-07	4.39e+06	0.000	1.000	1.000
Efectivo	-7.665e-07	1.79e-06	-0.429	0.668	-4.27e-0.6	2.74e-06
ar.L1	-0.9533	2.57e-11	-3.7e+10	0.000	-0.953	-0.953
ar.L2	-0.3232	1.58e-11	-2.04e+10	0.000	-0.323	-0.323
ma.L1	-0.7901	8.02e-12	-9.85e+10	0.000	-0.790	-0.790
ma.L2	-0.2707	1.72e-11	-1.57e+10	0.000	-0.271	-0.271
ma.L3	0.3025	4.38e-12	6.9e+10	0.000	0.302	0.302
Sigma2	5.451e-10	1.38e-10	3.947	0.000	2.74e-10	8.16e-10

Figura 19

Gráfico AUM Dato Real vs Predicción ARIMAX



Modelo VAR

Para la construcción del modelo predictivo, se seleccionaron inicialmente aquellas variables que presentaban un coeficiente de correlación con la variable objetivo (activos administrados por los FIC) superior a 0.5, dado que este umbral indica una relación moderada a fuerte. Este criterio permitió definir el primer modelo, enfocándose en las variables con mayor potencial explicativo.

Posteriormente, en el desarrollo del segundo modelo, se aplicó la prueba de causalidad o test de Wiener-Granger, la cual es usada para comprobar si los resultados de una variable sirven para predecir otra variable. Consiste en comparar una serie temporal X_t con una serie temporal Y_t bajo la siguiente afirmación:

Una serie temporal X_t causa otra serie temporal Y_t si los valores pasados de X_t ayudan a predecir los valores futuros de Y_t . Esto quiere decir que Y_t no sólo sería una función que depende de los valores pasados de Y_t , es decir, Y_{t-p} , sino que también dependería de los valores pasados de la serie temporal X_t , es decir, X_{t-p} , al aportar esta información relevante. (Acín Coello de Portugal, 2022)

Esta prueba permite identificar si una variable además de estar correlacionada con otra variable cause las alteraciones sobre esta, es decir, contribuye significativamente a predecir el comportamiento de otra en términos temporales. Para comprobar esto se realiza una prueba de hipótesis teniendo dos alternativas que son: H_0 : que no existe causalidad o alternativamente, H_a : que no se sepa que existe causalidad (Montero Granados, 2013). Para evaluar la relación de causalidad entre las variables seleccionadas según el primer modelo, se usó la función `grangercausalitytests` de `statmodels` (Statsmodels, 2024), de acuerdo a esto, las variables que no mostraron causalidad en el sentido de Granger fueron excluidas del modelo, al no aportar

información relevante para anticipar los cambios en la variable objetivo. Y este proceso (ver Figura 20) se repitió hasta quedar con las variables que, evaluada la causalidad, el resultado dictaba que contribuían significativamente. El modelo con mejor rendimiento fue el construido con las variables AUM FIC, Efectivo, CDT, Otros Depósitos a la Vista y Cuentas de Ahorro. La representación gráfica de la comparación entre las predicciones y el conjunto de prueba se muestra en la Figura 21, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

Figura 20

Diagrama de Flujo Procedimiento Modelo VAR

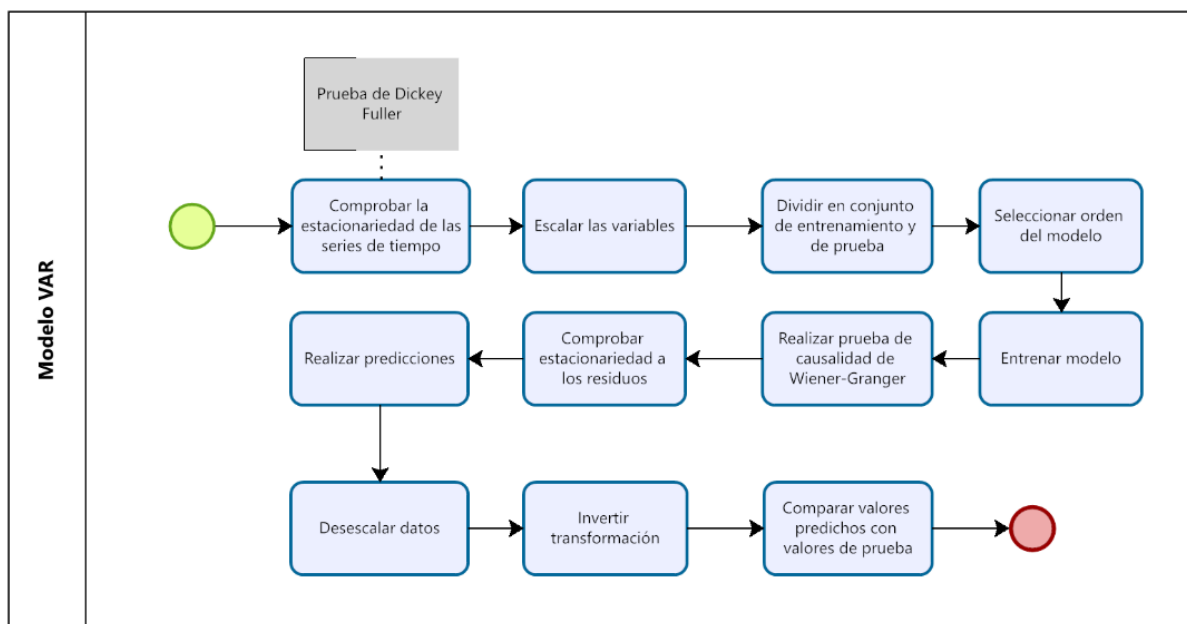
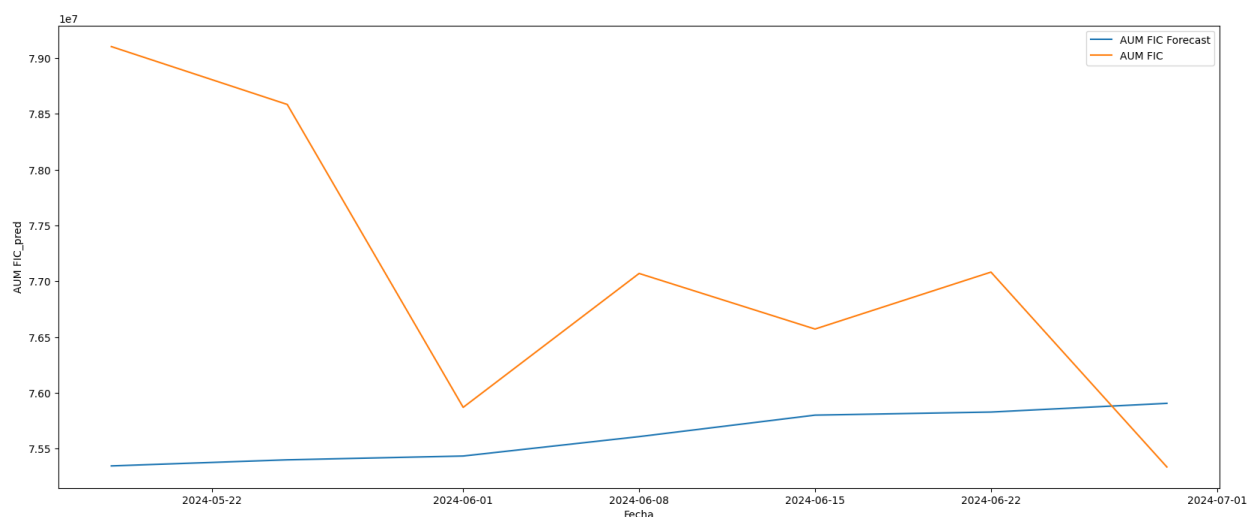


Figura 21

Gráfico AUM Dato Real vs Predicción VAR



Suavizado Exponencial

Se implementó un modelo de suavizamiento exponencial de Holt-Winters con componente multiplicativo, seleccionado por su capacidad para capturar la tendencia lineal y la estacionalidad presente en la serie temporal. La elección del enfoque multiplicativo tanto para la tendencia como para la estacionalidad se justifica por la presencia de patrones repetitivos regulares cuyo comportamiento varía proporcionalmente con el nivel de la serie.

Para el parámetro de periodo estacional, se evaluaron diferentes valores: 30, 60, 90, 120, 150 y 180 días, considerando los patrones mensuales observados en los datos como lo muestra la Figura 22. Tras comparar el rendimiento del modelo con cada configuración, se determina que un período de 180 días ofrece el mejor desempeño predictivo, reflejando de manera más precisa la dinámica estacional de la serie.

Figura 22

Descomposición Multiplicativa de la Serie por Año



Para la construcción del modelo se siguieron los pasos descritos en la Figura 23. La representación gráfica de comparación de la predicción del modelo y el conjunto de prueba se muestra en la Figura 24, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

Figura 23

Diagrama de Flujo Procedimiento Modelo Suavizado Exponencial

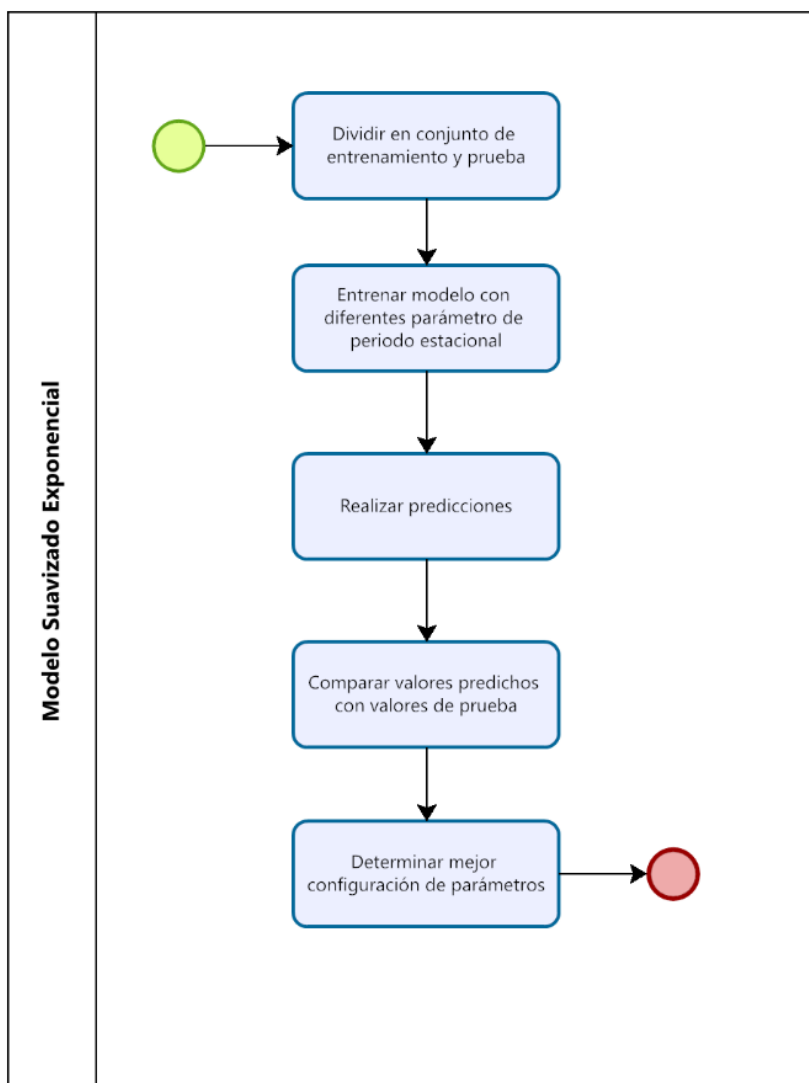
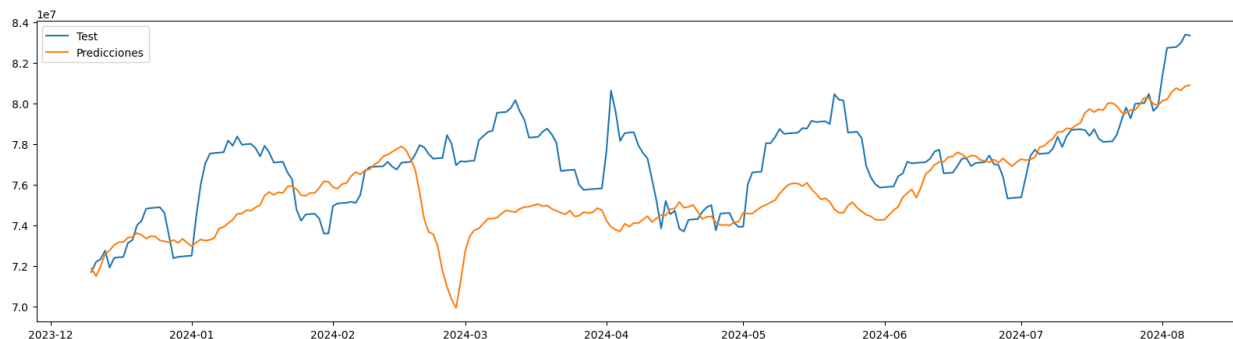


Figura 24

Gráfico AUM Dato Real vs Predicción Suavizado Exponencial



Modelo LSTM

Para optimizar el desempeño del modelo predictivo, se evaluaron cuatro configuraciones distintas, cada una con enfoques específicos en la estructura de los datos y el horizonte de predicción:

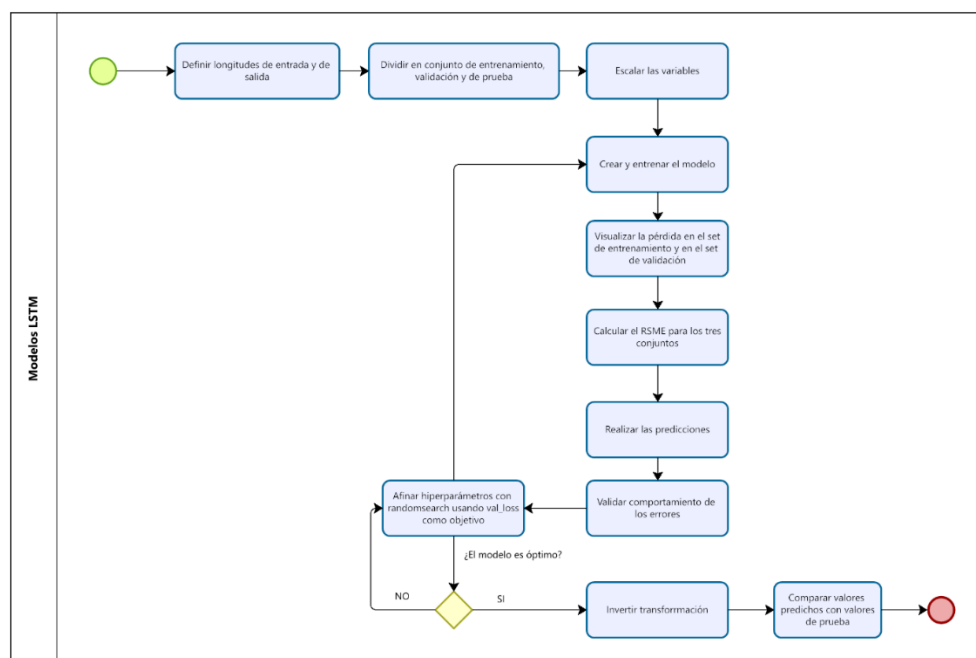
1. **Modelo Univariado - Unistep:** Utiliza únicamente la variable objetivo (AUM FIC) para realizar predicciones de un único paso hacia adelante a partir de los últimos 30 días (longitud de entrada = 30 y longitud de salida =1)
2. **Modelo Univariado - Multistep:** Basado solo en la variable objetivo (AUM FIC), pero orientado a predecir múltiples pasos futuros de forma secuencial a partir de los últimos 30 días (longitud de entrada = 30 y longitud de salida =3)
3. **Modelo Multivariado - Unistep:** Incorpora 8 variables exógenas (aquellas con coeficiente de correlación mayor a 0.5) junto con la variable objetivo (AUM FIC) para predecir un solo paso adelante a partir de los últimos 30 días (longitud de entrada = 30 y longitud de salida =1)

4. Modelo Multivariado - Unistep: Incorpora 2 variables exógenas (Efectivo y CDT) junto con la variable objetivo (AUM FIC) para predecir un solo paso adelante a partir de los últimos 30 días (longitud de entrada = 30 y longitud de salida = 1)

Cada configuración fue evaluada mediante métricas de desempeño como la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio (MAE). A partir de estos resultados, se seleccionaron las dos configuraciones con mejor rendimiento. Para la construcción de los modelos se siguieron los pasos descritos en la Figura 25.

Figura 25

Diagrama de Flujo Procedimiento Modelo LSTM



Los modelos que mejor rendimiento obtuvieron fueron el Univariado – Unistep y el Multivariado Unistep con dos variables, las métricas de evaluación de estos modelos se muestran en la Tabla 7. La representación gráfica de comparación de la predicción del modelo y el

conjunto de prueba se muestra en la Figura 26, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

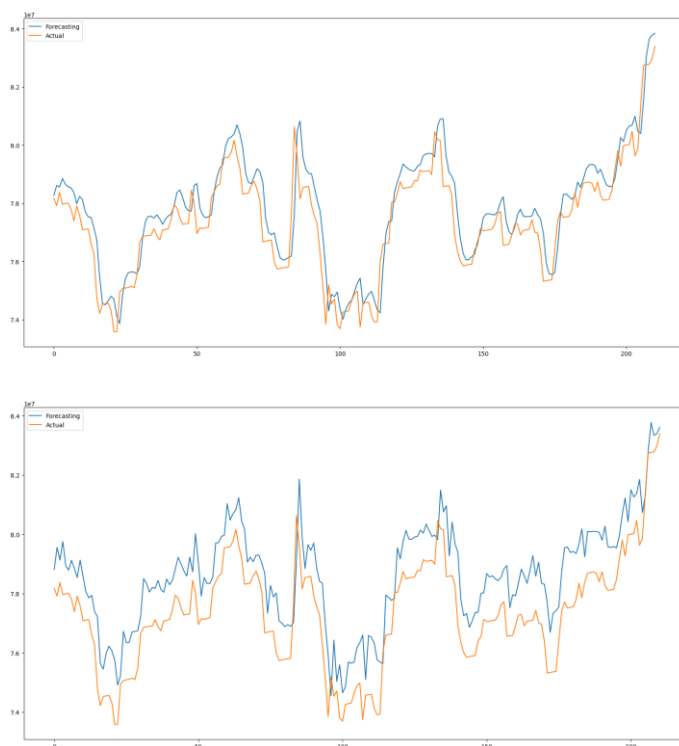
Tabla 7

Configuración Hiperparámetros Modelos LSTM

Modelo	Epochs	Batch Size	Units	Learning Rate	Val Loss
Univariado Unistep	150	256	128	0.05	0.02572
Multivariado					
Unistep (2 variables)	200	256	256	0.05	0.0334

Figura 26

Gráfico AUM Dato Real vs Predicción LSTM

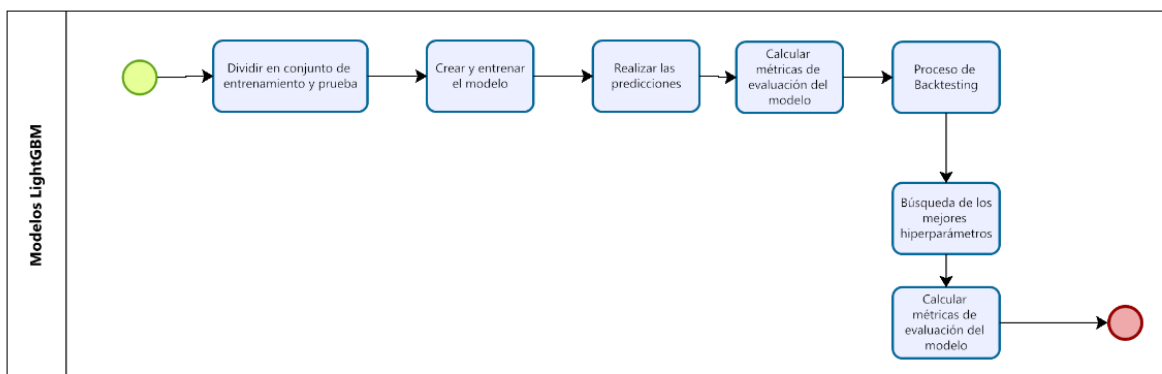


Nota. La Figura a) muestra la predicción del modelo LSTM univariado unistep y la imagen, la Figura b) muestra la predicción del modelo LSTM multivariado unistep.

LightGBM

Para este modelo se implementó inicialmente una configuración univariada, entrenando exclusivamente con la variable objetivo. Posteriormente, se aplicó un proceso de backtesting con reacondicionamiento y aumento progresivo del tamaño del conjunto de entrenamiento (Skforecast, s.f.), el cual inició con un 90% de los datos para el conjunto de entrenamiento (2.169 registros) y un 10% de los datos para el conjunto de prueba (242 registros) un enfoque especializado de validación cruzada de tipo Walk Forward en el cual se tiene una ventana deslizante en la que se agregan datos de nuevos al conjunto de entrenamiento manteniendo el orden temporal de los datos, mientras que el conjunto de prueba avanza se mantiene constante entrenamiento (Skforecast, s.f.) lo que permite entrenar frecuentemente adaptándose a las tendencias y patrones cambiantes y da una visión de cómo funcionará el modelo a futuro (Nagaraj, 2023). Luego, se realizó una optimización bayesiana de hiperparámetros, con el objetivo de mejorar el rendimiento del modelo al ajustar de forma eficiente los parámetros clave.

El mismo proceso de entrenamiento y validación se replicó incluyendo como variables exógenas aquellas que mostraron un coeficiente de correlación superior a 0.5 con los activos administrados de los FIC (AUM FIC). Finalmente, tras evaluar el desempeño de los modelos mediante métricas de precisión, se seleccionó la configuración que presentó el mejor rendimiento predictivo. Para la construcción de los modelos se siguieron los pasos descritos en la Figura 27.

Figura 27*Diagrama de Flujo Procedimiento Modelo LigthGBM*

Powered by
bizagi
Modeler

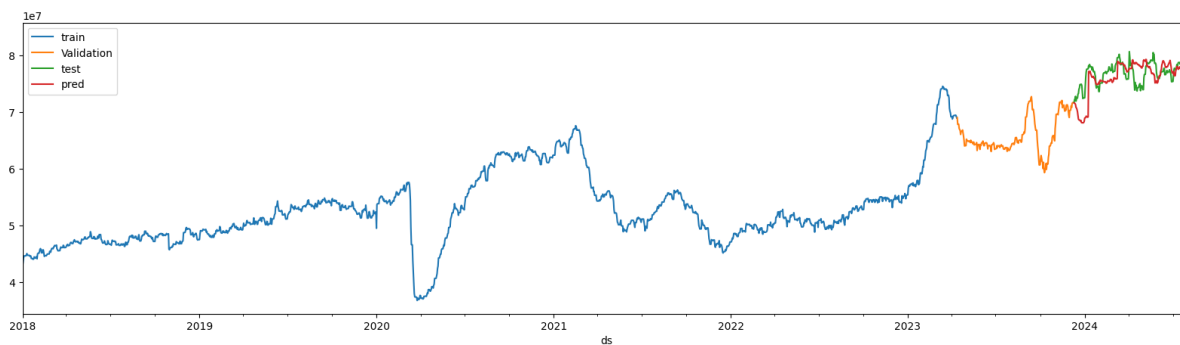
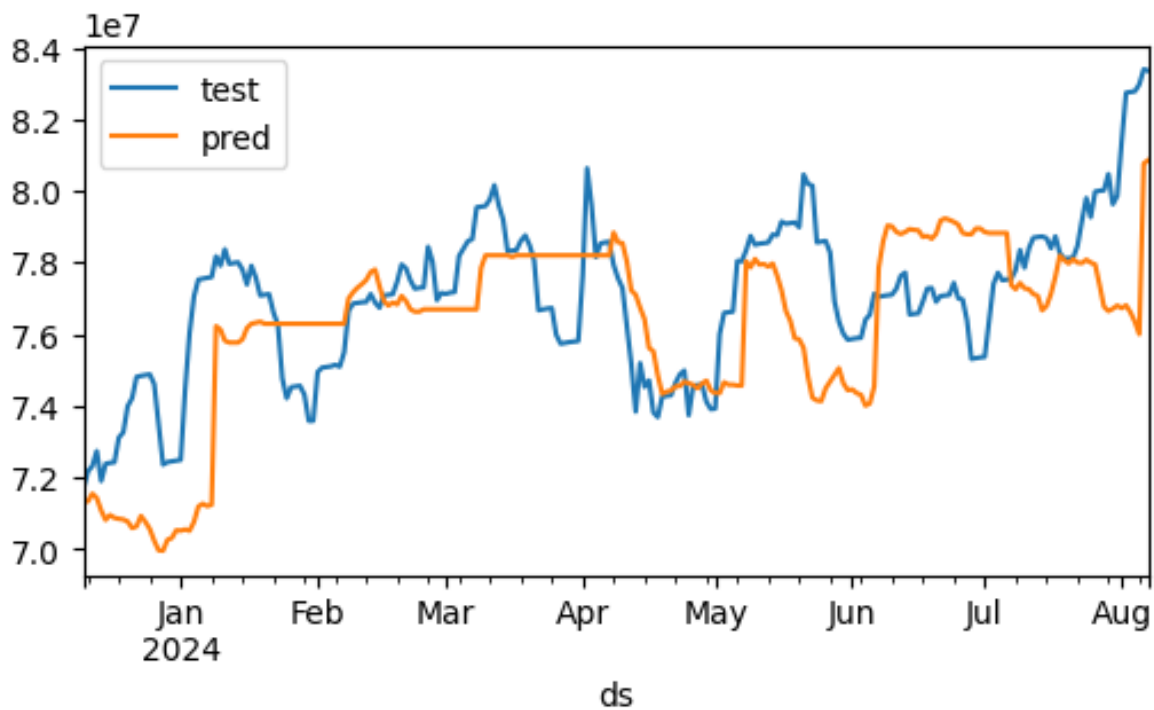
Los modelos que obtuvieron el mejor rendimiento fueron ajustados bajo la siguiente configuración:

- Modelo Univariado Autorregresivo con LGBM Regressor, este modelo utilizó las últimas 24 observaciones para generar predicciones. Su desempeño fue evaluado mediante un proceso de backtesting con reacondicionamiento y aumento progresivo del tamaño del conjunto de entrenamiento, realizando predicciones 30 días hacia adelante en cada iteración. La métrica utilizada para evaluar su rendimiento fue el Error Absoluto Medio (MAE);
- Modelo Multivariado Autorregresivo con LGBM Regressor, este modelo incorporó variables exógenas y fue configurado con 100 estimadores, una profundidad máxima de árbol de 3 y una tasa de aprendizaje de 0.01. Se entrenó utilizando las últimas 150 observaciones y su desempeño se evaluó bajo el mismo esquema de backtesting con reacondicionamiento y aumento progresivo del conjunto de entrenamiento, realizando también predicciones 30 días hacia adelante. La métrica de evaluación fue el MAE.

La representación gráfica de comparación de la predicción del modelo y el conjunto de prueba se muestra en la Figura 28, evidenciando el desempeño del modelo en la estimación del comportamiento de la serie temporal analizada.

Figura 28

Gráfico AUM Dato Real vs Predicción LightGBM



Nota. La Figura a) muestra la predicción del modelo LightGBM univariado y la Figura b) muestra la predicción del modelo LightGBM multivariado.

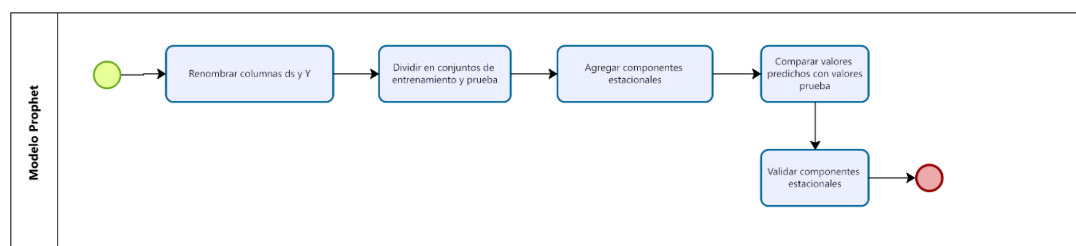
Prophet

El modelo fue parametrizado incorporando múltiples componentes de estacionalidad para capturar distintos patrones temporales en los datos. Se agregó una estacionalidad semanal (cada 7 días) para captar fluctuaciones recurrentes de corto plazo, una estacionalidad mensual (30 días) que permitió modelar de manera precisa los ciclos mensuales, y una estacionalidad anual (365.25 días) orientada a reflejar tendencias generales y comportamientos estacionales de largo plazo.

Esta misma configuración se aplicó al modelo que incluyó la variable exógena del índice COLTES, dado que esto mostró el mejor desempeño en términos predictivos. La incorporación de esta variable exógena permitió mejorar la capacidad del modelo para capturar dinámicas adicionales del mercado y optimizar la precisión de las predicciones sobre los activos administrados de los FIC. Para la construcción de los modelos se siguieron los pasos descritos en la Figura 29.

Figura 29

Diagrama de Flujo Procedimiento Determinación Modelo Prophet



Powered by

 Orange Modeler

El modelo que obtuvo un mejor rendimiento fue Prophet multivariado captando las tendencias semanales, mensuales y anuales como se muestra en la Figura 31. La representación gráfica de comparación de la predicción del modelo y el conjunto de prueba se muestra en la Figura 30.

Figura 30

Gráfico AUM Dato Real vs Predicción Prophet

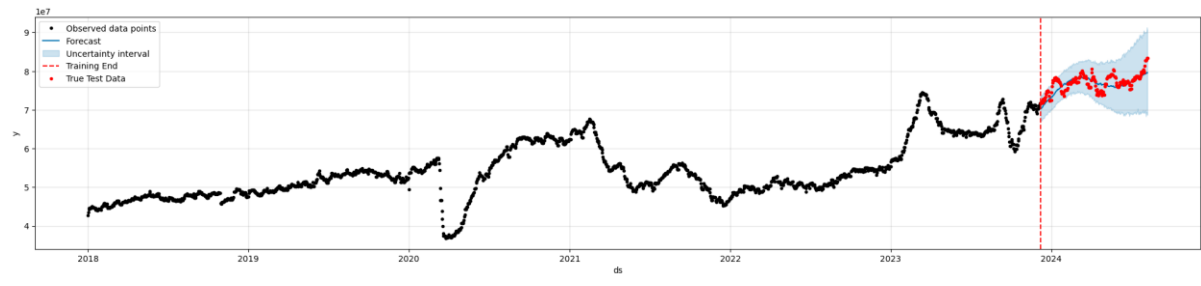
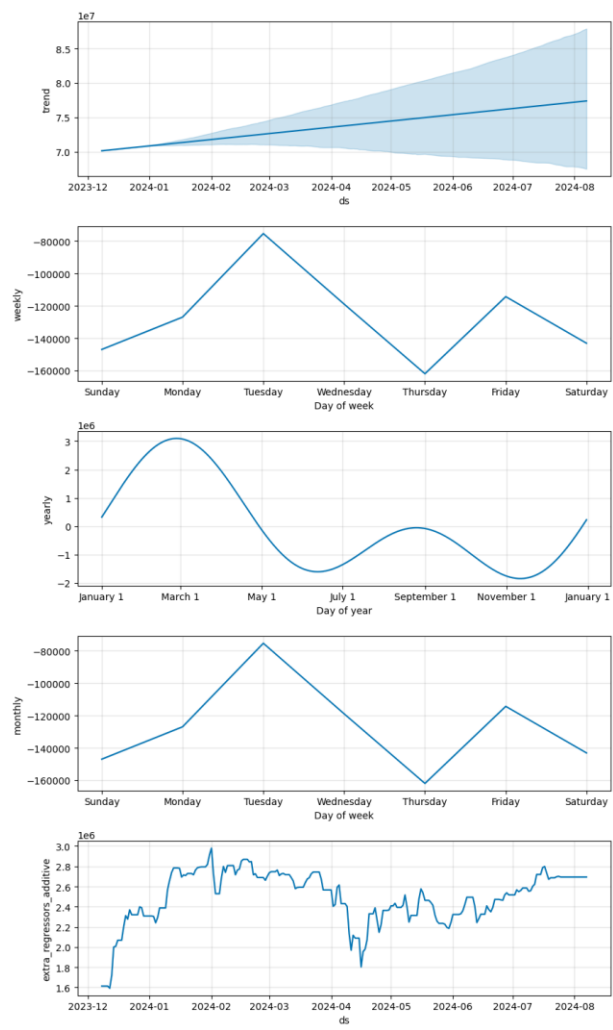


Figura 31

Tendencias Captadas por el Modelo Prophet Multivariado



Resultados

En este capítulo se presentan los resultados de los modelos con mejor rendimiento según las métricas de evaluación, seleccionando al menos un modelo representativo por cada método o técnica de predicción utilizada. Se analizan los resultados tanto desde una perspectiva numérica como desde el contexto de la realidad económica, destacando las contribuciones de cada modelo al conocimiento de la industria de los Fondos de Inversión Colectiva administrados por las sociedades fiduciarias. Asimismo, se examina la relación entre las variables identificadas y se establecen cuáles métodos pueden ser utilizados con mayor precisión para la predicción. Por último, se analiza cual es el aporte del modelo que presentó mejor rendimiento en comparación con el rendimiento de un modelo simple.

Resultados de la Evaluación

En esta etapa del estudio se realizó una comparación detallada de los modelos seleccionados, utilizando las métricas Raíz del Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE). Estas métricas permitieron evaluar objetivamente el desempeño de cada modelo en términos de precisión y ajuste a la relación temporal de la variable objetivo.

La selección de estas métricas responde a su uso frecuente en la evaluación de modelos predictivos (Chai y Draxler, 2014) debido su facilidad de interpretación, ya que se expresan en las mismas unidades que la variable objetivo. Esto permite que los errores sean explicados en la misma escala del AUM FIC, lo cual resulta clave para comunicar los resultados en un contexto no especializado en ciencia de datos, como se pretendía en este contexto. Además, el RMSE y el MAE tienen fundamentos geométricos bien establecidos: el RMSE se basa en la distancia euclidiana promedio y el MAE en la distancia Manhattan (Hodson, 2022).

Aún no existe consenso sobre los investigadores sobre cuál métrica es más adecuada, ya que algunos autores recomiendan el MAE por ser una medida más natural del error promedio (Willmott y Matsuura, 2005), mientras que otros prefieren el RMSE cuando la distribución de los errores es aproximadamente gaussiana (Chai y Draxler, 2014). Una solución intermedia ha sido proponer un promedio ponderado de ambas métricas o presentarlas juntas con sus respectivas ponderaciones. En este estudio, se decidió incluir ambas métricas considerando que el RMSE penaliza más los errores grandes, mientras que el MAE es más robusto ante valores atípicos (Cosio, 2021).

Se entrenaron y evaluaron un total de 18 modelos basados en diferentes algoritmos, de los cuales se seleccionaron los 9 modelos con mejor rendimiento para proceder con la comparación cuantitativa. La selección se fundamentó en la capacidad de los modelos para capturar las dinámicas temporales y minimizar los errores de predicción. En la Tabla 8 se presentan los resultados obtenidos en la evaluación comparativa.

Tabla 8

Comparación Métricas de Evaluación Modelos Predictivos AUM FIC

Modelo	RMSE	MAE
LSTM Multivariado	7.7 e+5	5.99 e +5
LSTM Univariado	7.97 e+5	6.30 e+5
Prophet Multivariado	1.80 e+6	1.44 e+6
VAR	2.03 e+6	1.63 e+6
ARIMAX	2.16 e+6	1.75 e+6
XGBoost	2.26 e+6	1.72 e+6
LightGBM	2.33 e+6	1.82 e+6
Suavizado Exponencial	2.55 e+6	1.94 e+6
ARIMA	3.92 e+6	3.81 e+6

Nota. El modelo LSTM multivariado unistep se parametrizó usando como variables exógenas CDT y Efectivo. El Modelo LSTM univariado es unistep. El Modelo Prophet multivariado se parametrizó usando como variable exógena el índice COLTES. El modelo VAR multivariado se parametrizó usando como variables exógenas Efectivo, CDT, Otros Depósitos a la vista y Cuentas de Ahorro. El modelo ARIMAX se parametrizó usando como variable exógena Efectivo.

Análisis de Resultados

La evaluación de los modelos predictivos resalta que el LSTM multivariado es la técnica predictiva con mejor desempeño para pronosticar el comportamiento del AUM de los FIC de Renta Fija nacional para entidades públicas, Renta Fija pesos de liquidez y Renta Fija pesos de corto plazo. Este resultado evidencia que los modelos de aprendizaje profundo capturan de manera más efectiva las dinámicas complejas y no lineales de estas series temporales, superando a los modelos econométricos y los modelos de boosting como XGBoost y LightGBM.

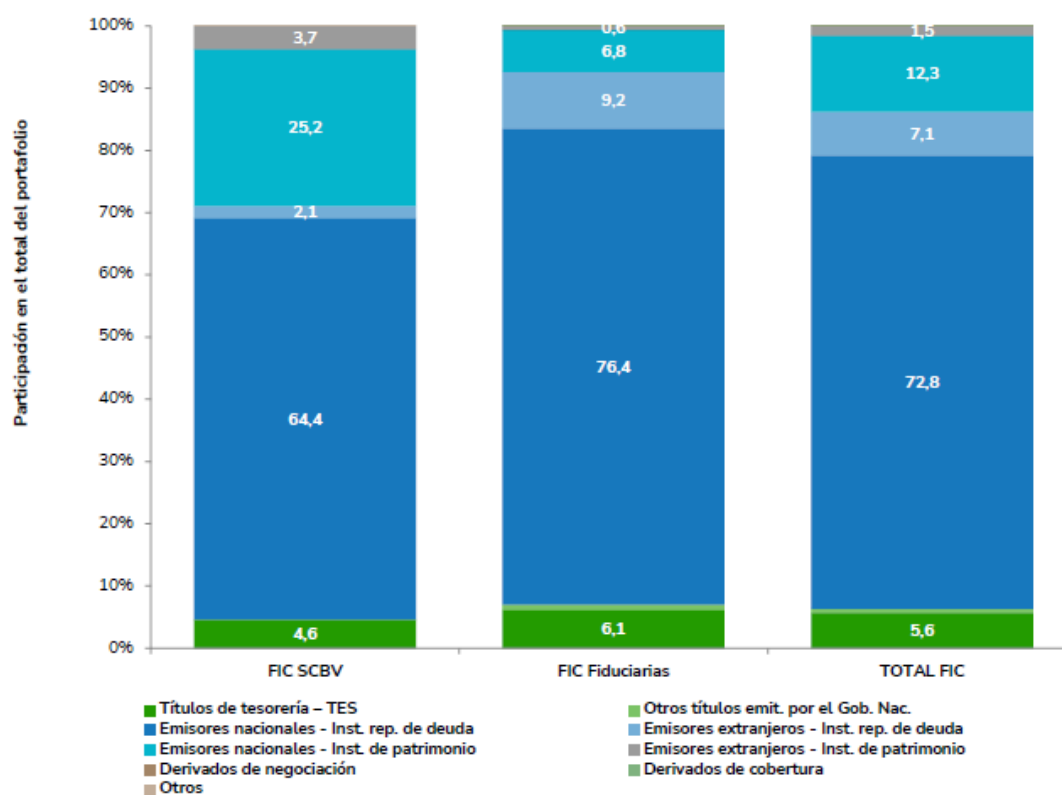
Asimismo, el análisis revela que los modelos multivariados dominan los primeros puestos de desempeño, ocupando 4 de los 5 mejores lugares. Modelos como el LSTM multivariado, Prophet multivariado, VAR y ARIMAX demuestran la relevancia de incorporar variables exógenas para mejorar la precisión del pronóstico. Este resultado confirma la importancia de considerar factores externos que influyen directamente en la evolución de los activos administrados.

Entre las variables predictoras, destacan el Efectivo y los CDT, componentes clave del agregado monetario M1, como los principales impulsores del comportamiento de los FIC. Esto se debe a dos factores fundamentales: Naturaleza de Inversión en Renta Fija, los FIC analizados se enfocan en inversiones en instrumentos de deuda emitidos por entidades nacionales, siendo los

CDT uno de los instrumentos más representativos de esta categoría, como lo muestra la Figura 32. Su comportamiento impacta directamente el desempeño de los fondos; y al ser instrumentos de inversión, los FIC están intrínsecamente ligados al efectivo, que representa la disponibilidad de recursos para inversión. Los cambios en el agregado monetario M1 afectan el consumo y la inversión, influyendo así los movimientos en los activos administrados. Estos hallazgos demuestran cómo el análisis de agregados monetarios puede guiar estrategias de inversión más efectivas, permitiendo a los agentes económicos adaptarse proactivamente a los ciclos económicos (EBC Financial Group, 2024).

Figura 32

Composición del Portafolio de Inversión de los FIC



Nota. Tomado de (Superintendencia Financiera de Colombia, 2024a)

Aunque los modelos de aprendizaje profundo como el LSTM multivariado lograron las mejores métricas de predicción, el desempeño destacado del modelo VAR (Vector Autoregressive) es especialmente relevante. Este modelo no solo superó a los algoritmos de aprendizaje automático basados en Boosting (XGBoost y LightGBM), sino que también se posicionó como el mejor modelo econométrico autorregresivo.

Este resultado es significativo por varias razones: primero, el modelo VAR exige considerablemente menos recursos computacionales que los modelos de aprendizaje automático y profundo. Esto lo convierte en una opción viable para entornos con limitaciones tecnológicas. Segundo, la configuración de un modelo VAR es más sencilla en comparación con la complejidad de ajuste y optimización de hiperparámetros en modelos de aprendizaje automático. Tercero, permite realizar análisis de causalidad, como la prueba de Granger, que facilita entender cómo unas variables impactan directamente a otras. Esta capacidad interpretativa es clave para la toma de decisiones fundamentadas. Y cuarto, gracias a su bajo requerimiento computacional y facilidad de implementación, el modelo VAR es ideal para organizaciones o entornos con infraestructura tecnológica básica. Esto lo posiciona como un punto de partida efectivo para desarrollar análisis predictivos sin necesidad de herramientas avanzadas.

Evaluación del Desempeño del Modelo LSTM Multivariado

Un aspecto clave del análisis es el desempeño del modelo LSTM multivariado, que presentó un RMSE de 7.7×10^5 . Esto implica que la diferencia promedio entre el valor real del AUM de los FIC y el valor predicho por este modelo es de aproximadamente \$770.000 millones. Si se considera un valor estimado del AUM de \$80 billones, este error representa un margen de aproximadamente 1%. En términos absolutos, esto significa que las predicciones del modelo se encuentran dentro de un rango entre \$79,23 billones y \$80,77 billones.

Para validar la calidad del RMSE obtenido por el modelo LSTM multivariado, se utilizaron dos criterios comparativos adicionales. Primero la comparación con la desviación estándar de la serie, La desviación estándar del AUM FIC es de 9.7×10^6 , mientras que el RMSE del modelo es de 7.7×10^5 . Esto indica que el error promedio del modelo es significativamente menor que la variabilidad natural de la serie, lo que sugiere que el modelo es capaz de capturar y explicar parte importante de la variabilidad de los datos.

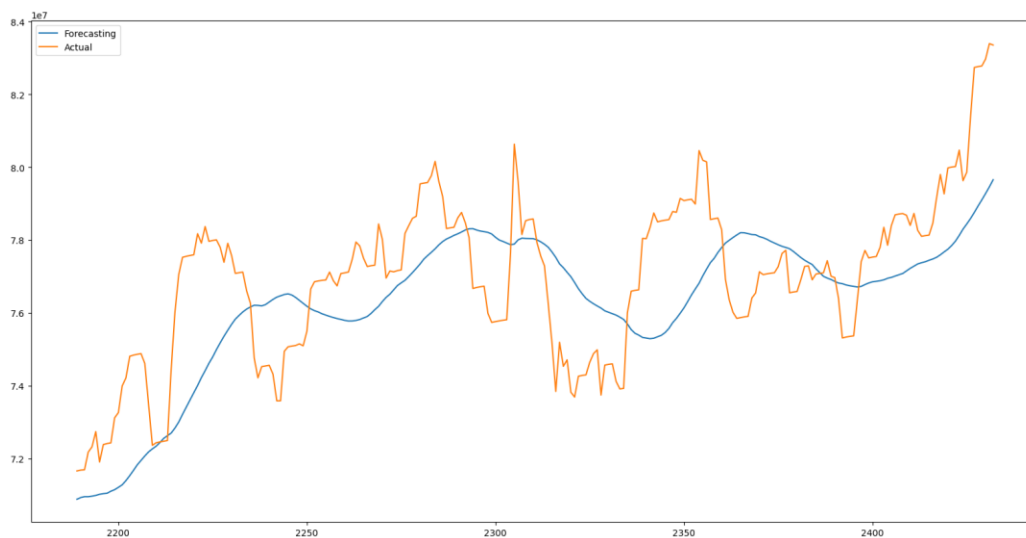
El segundo criterio es la comparación con un modelo simple como el de media móvil, el cual presentó un RMSE de 1.96×10^6 , en la Figura 33 se muestra la predicción con este modelo. Comparado con el LSTM, se observa una mejora del 61% en la predicción:

$$Mejora = \left[\left(\frac{7.7 \times 10^5}{1.96 \times 10^6} \right) - 1 \right] * 100 \approx 61 \%$$

Este resultado refuerza la efectividad del modelo LSTM frente a enfoques más simples, demostrando que es mucho más preciso al capturar las dinámicas temporales del AUM de los FIC.

Figura 33

Gráfico AUM Dato Real vs Predicción Media Móvil



Conclusiones

Los principales hallazgos de este estudio destacan la importancia de ciertas variables clave en la predicción del AUM de los FIC. El análisis del coeficiente de valoración revela que factores como el Efectivo, los CDT y otros depósitos a la vista mantienen una relación significativa y positiva con los activos administrados, sugiriendo que un incremento en estas variables podría estar asociado con el crecimiento de los FIC. Esta relación subraya la relevancia de estos elementos en la dinámica de inversión y en la toma de decisiones dentro del sector financiero.

En términos de modelado, el LSTM multivariado se consolidó como la mejor alternativa para la predicción del AUM de los FIC de Renta Fija nacional, superando tanto a los modelos econométricos como a los algoritmos de boosting. Su capacidad para capturar patrones complejos y no lineales en los datos financieros lo convierte en una herramienta poderosa para generar pronósticos más precisos y adaptables a las fluctuaciones del mercado. Además, la inclusión de variables exógenas, como el Efectivo y los CDT, pertenecientes al agregado monetario M1, jugó un papel determinante en la mejora de la precisión de los modelos predictivos. Estos factores reflejan la estrecha relación entre la liquidez del mercado y el comportamiento de los fondos de inversión colectiva, resaltando la necesidad de un enfoque integral que combine técnicas de modelado avanzado con un análisis macroeconómico sólido.

El modelo LSTM multivariado demostró un alto nivel de precisión, con un RMSE de 7.7×10^5 , lo que equivale a un margen de error del 1% respecto al valor estimado del AUM de los FIC. La comparación con la desviación estándar de la serie indica que el modelo captura eficazmente la variabilidad de los datos, proporcionando predicciones confiables dentro de un rango aceptable. En términos de desempeño comparativo, el modelo LSTM mostró una mejora

del 61% en la precisión de la predicción en comparación con un modelo de medias móviles, reafirmando su capacidad superior para modelar las dinámicas temporales del AUM de los FIC. Este resultado resalta la ventaja de los enfoques basados en el aprendizaje profundo sobre modelos más simples, consolidando su utilidad en escenarios donde la alta precisión en la toma de decisiones financieras es crucial.

Si bien el multivariado LSTM obtuvo las mejores métricas de predicción, el modelo VAR se destacó como una alternativa altamente efectiva, logrando superar a los algoritmos de boosting y posicionándose como el mejor modelo econométrico autorregresivo. Su capacidad para ofrecer resultados competitivos con menores requisitos computacionales y su facilidad de configuración lo hacen una opción atractiva para entornos con recursos tecnológicos limitados. Además, el modelo VAR proporciona ventajas analíticas clave, como la posibilidad de realizar análisis de causalidad mediante la prueba de Granger, lo que permite a los tomadores de decisiones comprender mejor las interacciones entre variables y desarrollar estrategias informadas sin la complejidad propia de los modelos de aprendizaje automático avanzados.

Recomendaciones

A pesar de que los modelos de aprendizaje automático ofrecen una alta capacidad predictiva, su naturaleza de "caja negra" limita la interpretabilidad de los factores que explican sus resultados. Para mitigar esta limitación, se sugiere la aplicación para trabajos futuros de técnicas como SHAP (SHapley Additive exPlanations) o LIME (Local Interpretable Model-agnostic Explanations), que permiten generar explicaciones claras y comprensibles sobre el impacto de cada variable en las predicciones del modelo. Estas herramientas no solo mejoran la transparencia del proceso de modelado, sino que también brindan una visión más detallada de los factores que afectan el comportamiento de los FIC, lo cual es esencial para la toma de decisiones informadas.

El presente trabajo se desarrolló utilizando una ventana temporal de 2018 a 2024, lo que permitió obtener resultados significativos para el análisis de los Fondos de Inversión Colectiva (FIC). Sin embargo, dado que la serie histórica de datos está disponible desde 2009, sería valioso replicar este ejercicio utilizando una ventana temporal más amplia, que abarque desde 2009 hasta 2024. Este enfoque permitiría comparar si el uso de una mayor cantidad de datos históricos mejora el rendimiento de los modelos de predicción. La extensión de la ventana temporal podría proporcionar una visión más completa de las tendencias a largo plazo y permitir a los modelos captar patrones más profundos y representativos de la dinámica económica y financiera. Además, este ejercicio comparativo ayudaría a evaluar la estabilidad y robustez de los modelos a lo largo del tiempo, lo cual es esencial para determinar su aplicabilidad en escenarios futuros.

Aunque la selección de variables en este estudio se llevó a cabo con base en el criterio de los expertos en inversiones consultados, para un ejercicio futuro se sugiere considerar un enfoque estadístico para la selección de variables, como el Análisis de Componentes Principales (PCA,

por sus siglas en inglés). El PCA permite reducir la dimensionalidad del conjunto de datos, extrayendo las componentes principales que explican la mayor parte de la varianza en las variables originales. Este enfoque podría ofrecer una selección más objetiva y cuantitativa de las variables más relevantes para la predicción de los activos administrados por los FIC, mejorando la eficiencia del modelo y reduciendo el riesgo de sobreajuste. Además, el uso de PCA podría revelar relaciones subyacentes entre las variables que no son evidentes a través de un análisis puramente subjetivo, lo que enriquecería el modelo y lo haría más robusto a cambios en las condiciones del mercado.

Para trabajos futuros, se recomienda incluir la métrica de R^2 ajustada dentro de las métricas de evaluación de los modelos. Esta métrica penaliza la inclusión de predictores que no aportan significativamente a la predicción, permitiendo una mejor evaluación de la calidad del modelo. Además, podría servir como un indicador del valor de las variables exógenas incorporadas en el análisis, ayudando a determinar su verdadera relevancia en la predicción del AUM de los FIC.

Referencias

- Acín Coello de Portugal, C. (2022). *La Causalidad de Granger en el análisis y la previsión de series temporales clásicas, de intervalo y de historiograma. Aplicación de mercados financieros* [Trabajo Fin de Grado, Universidad Pontificia Comillas].
<https://repositorio.comillas.edu/xmlui/handle/11531/62362>
- Ampountolas, A. y Legg, M. (2024). Predicting daily hotel occupancy: A practical application for independent hotels. *Journal of Revenue and Pricing Management*, 23(3), 197-205.
<https://doi.org/10.1057/s41272-023-00445-7>
- Asociación de Fiduciarias de Colombia. (2016). *Estudio sobre Fondos de Inversión Colectiva en Colombia: Evolución y Oportunidades*. ANIF. <https://www.asofiduciarias.org.co/wp-content/uploads/2017/10/estudios-fondos-inversion-colectiva-asofiduciarias.pdf>
- Asociación de Fiduciarias de Colombia. (2022). *Primer Informe de Sostenibilidad del Sector Fiduciario* (No. 1; p. 106). <https://www.asofiduciarias.org.co/wp-content/uploads/2022/11/Primer-Informe-de-Sostenibilidad-del-Sector-Fiduciario.pdf>
- Asociación de fiduciarias de Colombia. (2024). *Estadísticas del Sector*.
<https://www.asofiduciarias.org.co/activos-y-negocios-administrados-por-tipologia/>
- Banco de la República de Colombia. (2020). *Ficha Metodológica Agregados Monetarios y crediticios*. <https://www.banrep.gov.co/sites/default/files/ficha-metodologica-agregados-y-cartera.pdf>
- Banco de la República de Colombia. (12 de abril de 2021). *Ficha Metodológica Tasas de interés de captación y operaciones del mercado monetario*.
<https://www.banrep.gov.co/sites/default/files/ficha-metodologica-tasas-de-captacion.pdf>

Banco de la República de Colombia. (2024a). *Estadísticas económicas*.

<https://suameca.banrep.gov.co/estadisticas-economicas/#!/informacionSerie/59/Tasas%20de%20inter%C3%A9s%20de%20pol%C3%ADtica%20monetaria>

Banco de la República de Colombia. (2024b). *Estadísticas económicas*.

<https://suameca.banrep.gov.co/estadisticas-economicas/#!/informacionSerie/230001/Agregados%20monetarios%20y%20crediticios/M1,%20M2,%20M3%20y%20base%20monetaria,%20semanal>

Banco de la República de Colombia. (2024c). *Estadísticas económicas*.

<https://suameca.banrep.gov.co/estadisticas-economicas/#!/informacionSerie/220003/Tasas%20de%20inter%C3%A9s/CDT's%20a%2090%20%28DTF%29,%20180%20y%20360%20d%C3%ADas>

Barbará, Y. (2021). *Apuntes de cátedra: Elaboración del árbol de problemas y de objetivos*

[Apuntes de cátedra, Facultad de Ciencia Política y Relaciones Internacionales Universidad Nacional de Rosario]. <http://hdl.handle.net/2133/23106>

Bermejo Iglesias, A., Torres Pruñonosa, J. y Masferrer Llavínés, N. (2009). *La utilización de los modelos ARIMA en la estimación del precio de acciones: Una aplicación preliminar.*

Administrando en entornos inciertos = managing in uncertain environment, 2009, ISBN 978-84-7356-609-4, 99. <https://dialnet.unirioja.es/servlet/articulo?codigo=3118810>

Bolsa de Valores de Colombia. (2020). *Metodología familia de índices COLTES.*

<https://media.graphassets.com/8hgOIMlrTaiJ6NBJDm4Q>

Bolsa de Valores de Colombia. (2024a). *Divisas.* <https://bvc.com.co/>,

https://bvc.com.co//mercado-local-en-linea?tab=divisas_mercado-spot

- Bolsa de Valores de Colombia. (2024b). *Índices Renta Fija*. <https://bvc.com.co/>,
https://bvc.com.co/mercado-local-en-linea?tab=divisas_mercado-spot
- Bolsa de Valores de Colombia. (2024c). *Renta Fija*. <https://www.bvc.com.co/renta-fija-descripcion-general>
- Burnham, K. P., y Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, 28(2), 111-119.
<https://doi.org/10.1071/wr99107>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Codificando Bits. (julio 20 de 2019). *¿Qué son las Redes LSTM?*.
<https://codificandobits.com/blog/redes-lstm/>
- Colombia. (2010). *Decreto 2555 de 2010*. Departamento Administrativo de la Función Pública.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=40032>
- Cosio, N. A. L. (diciembre 21 de 2021). *Métricas en regresión*. Medium.
<https://medium.com/@nicolasarrioja/m%C3%A9tricas-en-regresi%C3%B3n-5e5d4259430b>
- Departamento Nacional de Estadística. (2019). *FICHA METODOLÓGICA ÍNDICE DE PRECIOS AL CONSUMIDOR IPC*.
<https://www.dane.gov.co/files/investigaciones/fichas/precios-y-costos/DSO-IPC-FME-001-V9.pdf>

Departamento Nacional de Estadística. (2024). *Índice de Precios al Consumidor (IPC)*.

<https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc>

Das, N., Kish, R. y Muething, D. (2005). Modeling Hedge Fund Returns. *Financial Decisions*, Article 4, Fall.

https://www.researchgate.net/publication/266161662_Modeling_Hedge_Fund_Returns

Datos Abiertos de Colombia. (2024). *Rentabilidades de los Fondos de Inversión Colectiva (FIC)*.

https://www.datos.gov.co/Hacienda-y-Credito-Publico/Rentabilidades-de-los-Fondos-de-Inversion-Colectiva/qhpu-8ixx/about_data

Decreto 1242 de 2013. Por el cual se sustituye la Parte 3 del Decreto número 2555 de 2010 en lo relacionado con la administración y gestión de los Fondos de Inversión Colectiva. 14 de junio de 2013.

<https://www.alcaldiabogota.gov.co/sisjur/normas/Normal.jsp?i=127208&dt=S>

De la Fuente Fernández, S. (s.f.). *SERIES TEMPORALES, MODELO ARIMA*. Universidad

Autónoma de Madrid. <https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>

EBC Financial Group. (17 de julio de 2024). *El significado del dinero M1 y su uso de datos*.

<https://www.ebc.com/es/forex/70163.html>

Fourkiotis, K. P., y Tsadiras, A. (2024). Applying Machine Learning and Statistical Forecasting

Methods for Enhancing Pharmaceutical Sales Predictions. *Forecasting*, 6(1), 170-186.

<https://doi.org/10.3390/forecast6010010>

- Ghosh, S., y Ranjan, A. (2023). A machine learning approach to gdp nowcasting:an emerging market experience. *Bulletin of Monetary Economics and Banking*, 26(0), 33-54.
<https://doi.org/10.59091/1410-8046.2055>
- Groene, N., y Zakharov, S. (2024). Introduction of AI-based sales forecasting: How to drive digital transformation in food and beverage outlets. *Discover Artificial Intelligence*, 4(1), 1. <https://doi.org/10.1007/s44163-023-00097-x>
- Grupo Proindustria. (2024, julio 11). *Análisis de los dos años del gobierno de Petro*.
<https://grupoproindustria.org/analisis-de-los-dos-anos-del-gobierno-de-petro/>
- Herrera, D., y Angarita, G. (2023). Política tributaria en el gobierno de Iván Duque: Del impulso al crecimiento a corregir las finanzas públicas tras la pandemia. En J. Julio Estrada (Ed.), *Seguimiento y análisis de políticas públicas en Colombia 2022* (p 11-33). Universidad Externado de Colombia. <https://bdigital.uexternado.edu.co/handle/001/13430>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481-5487.
<https://doi.org/10.5194/gmd-15-5481-2022>
- Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., Ye, Q., y Liu, T.-Y. (4 de diciembre de 2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. Advances in Neural Information Processing Systems 30 (NIP 2017). <https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/>
- Korstanje, J. (2021). *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*. Apress. <https://doi.org/10.1007/978-1-4842-7150-6>

- Kvam, P., y Vidakovic, B. (2007). *Nonparametric Statistics with Applications to Science and Engineering*. John Wiley & Sons. <https://maktab-sms.ir/Uploads/Ebooks/9339037c-8955-4792-afb7-b003c6df9ce7.pdf>
- Lizama, P., Boccardo, G., Díaz, I., García, C., León, M., Ruiz, F., y Torres, F. (2014, noviembre). *Guía de Asociación entre variables (Pearson y Spearman en SPSS)*. Universidad de Chile. https://www.u-cursos.cl/facso/2014/2/SO01007/1/material_docente/bajar?id_material=994690
- LVA Colombia. (2024). *Estadísticas*. <https://sificcolombia.lvaindices.com/estadisticas/>
- Mahfooz, A., y Phillips, J. L. (2024). Conditional Forecasting of Bitcoin Prices Using Exogenous Variables. *IEEE Access*, 12, 44510-44526. <https://doi.org/10.1109/ACCESS.2024.3381516>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048-3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Mejía Corredor, J. F., y Mejía Corredor, A. A. (2020). *Retos de la industria de fondos de inversión colectiva en Colombia frente a las tendencias globales de gestión y distribución* [Trabajo Maestría, Universidad EAFIT]. <http://hdl.handle.net/10784/17420>
- Mertz, D. (2021). *Cleaning Data for Effective Data Science: Doing the Other 80% of the Work with Python, R, and Command-line Tools* (1.ª ed.). Pack Publishing. <https://viewer-ebscohost-com.bibliotecavirtual.unad.edu.co/EbscoViewerService/ebook?an=2902696&callbackUrl=https%3a%2f%2fresearch.ebsco.com&db=e000xww&format=EB&profId=eds&lpId=lp>

[Cover-2&ppid=pp_FM-1&lang=es&location=edm&isPLink=False&requestContext=&profileIdentifier=qcagk4&recordId=xmvqmh75bn](#)

Momparler, A., Carmona, P., y Climent, F. (2024). Catalyzing Sustainable Investment: Revealing ESG Power in Predicting Fund Performance with Machine Learning. *Computational Economics*. <https://doi.org/10.1007/s10614-024-10618-0>

Montero Granados, R. (2013, marzo). *Test de causalidad*. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. <https://www.studocu.com/co/document/universidad-del-cauca/contabilidad-y-finanzas-para-emprendedores/montero-r-2013-test-de-causalidad-documentos-de-trabajo-en-economia-aplicada-universidad-de-granada-espana/60652690>

Mora Caballero, M. (19 de octubre de 2023). *Tiempedia: Modelos de Suavizamiento Exponencial*. LinkedIn. <https://es.linkedin.com/pulse/tiempedia-modelos-de-suavizamiento-exponencial-mora-caballero-pdn1e>

Natural History Museum. (s.f.). *Past 5—The Past of the Future* -. Recuperado 14 de enero de 2025, de <https://www.nhm.uio.no/english/research/resources/past/index.html>

Novales, A. (2017, noviembre). *Modelos vectoriales autoregresivos (VAR)*. Universidad Complutense. https://www.ucm.es/data/cont/media/www/pag-41459/VAR_new.pdf

Ortega Álzate, J. D., y Quirama Hernández, S. M. (2020). *Evolución de los fondos de inversión colectiva en Colombia 2009—2016* [Trabajo Pregrado, Tecnológico de Antioquia Institución Universitaria]. <https://dspace.tdea.edu.co/handle/tdea/674>

Pal, A. y Prakash, P. (2017). *Practical Time Series Analysis: Step by Step Guide Filled with Real World Practical Examples*. Packt Publishing.

- Panchal, S. A., Ferdouse, L., y Sultana, A. (2024). *Comparative Analysis of ARIMA and LSTM Models for Stock Price Prediction*. 2024 IEEE/ACIS 27th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 240-244. <https://doi.org/10.1109/SNPD61259.2024.10673919>
- Pandas. (2024a). *pandas.DataFrame.resample—Documentación de pandas 2.2.3*. Pandas. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.resample.html>
- Pandas. (2024b). *pandas.to_datetime—Documentación de pandas 2.2.3*. https://pandas.pydata.org/docs/reference/api/pandas.to_datetime.html
- Ramírez-Córdoba, G. L. (2014). La inversión colectiva en Colombia: Caracterización y análisis. *Revista Soluciones de Posgrado EIA*, (9), 13-37. <https://repository.eia.edu.co/entities/publication/4b9bb51e-a868-42a9-a0c9-e396890c5908>
- Rayón, E. L. (2015). Modelo multifactor para analizar la exposición de los hedge funds a factores de riesgo macroeconómico (Multifactor Model for the Analysis of the Vulnerability of Hedge Funds to Macroeconomic Risk Factors). *Economía teoría y práctica*, Article 42. <https://doi.org/10.24275/ETYP/AM/NE/422015/Leyva>
- Rendón-González, A. M. (2021). Indicador de alerta temprana para la detección de vulnerabilidades de los fondos de inversión colectiva (FIC). *ODEON*, 19, Article 19. <https://doi.org/10.18601/17941113.n19.03>
- Salgado Gaspar, A. (2024). *Predicción de saldos de personas jurídicas de un fondo de inversión de corto plazo* [Tesis de Maestría, Universidad de las Américas]. <http://dspace.udla.edu.ec/handle/33000/15849>

- Santamarina Campos, V. (2024). *Análisis de problemas y objetivos en trabajos finales de grado, máster y tesis doctorales utilizando el enfoque del marco lógico*. Universitat Politècnica de Valencia. <https://riunet.upv.es/handle/10251/205426>
- Sargent, T. J. (1988). *Teoría macroeconómica*. Antoni Bosch editor.
- Shabbir, N., Ahmadiyahangar, R., Rosin, A., Jawad, M., Kilter, J., y Martins, J. (2023). *XgBoost based Short-term Electrical Load Forecasting Considering Trends & Periodicity in Historical Data*. 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), 1-6. <https://doi.org/10.1109/ETFG55873.2023.10407926>
- SIFIC. (2021, febrero 21). *Metodología categorización fondos de inversión colectiva*. <https://sificolombia.com/wp-content/uploads/2021/04/20210216-Metodologi%CC%81a-Categorizacio%CC%81n-FICs-Versio%CC%81n-6.0-Final.pdf>
- SixSigma. (2024, julio 31). *Anderson-Darling Normality Test: A Complete Guide*. <https://www.6sigma.us/six-sigma-in-focus/anderson-darling-normality-test/>
- Skforecast. (s.f.). *Pronosticador de backtesting—Skforecast Docs*. Recuperado el 15 de enero de 2025, de https://skforecast.org/0.14.0/user_guides/backtesting.html
- Statsmodels. (2024). *statsmodels.tsa.stattools.grangercausalitytests—Statsmodels 0.15.0 (+605)*. <https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.grangercausalitytests.html>
- Sukhobok, D., Nikolov, N., y Roman, D. (2017). *Tabular Data Anomaly Patterns*. International Conference on Big Data Innovations and Applications (Innovate-Data), 25-34. <https://doi.org/10.1109/Innovate-Data.2017.10>

Superintendencia Financiera de Colombia. (2012, octubre 1). *Acerca de la SFC*.

<https://www.superfinanciera.gov.co/publicaciones/60607/nuestra-entidadacerca-de-la-sfc-60607/>

Superintendencia Financiera de Colombia. (2024a). *Informe Actualidad del Sistema Financiero Colombiano* (p. 24).

<https://www.superfinanciera.gov.co/publicaciones/10115426/resultados-del-sistema-financiero-colombiano-octubre-de-2024/>

Superintendencia Financiera de Colombia. (2024b). *Rentabilidades FIC*.

<https://www.superfinanciera.gov.co/ReporteInformacionFIC/>

Tapia, C. E. F., y Cevallos, K. L. F. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-darling, ryan-joiner, shapiro-wilk y kolmogórov-smirnov. *Societas*, 23(2), 83-106.

<https://portal.amelica.org/ameli/jatsRepo/341/3412237018/index.html>

Taylor, S. J., y Letham, B. (2017). *Forecasting at scale*. PeerJ Inc.

<https://doi.org/10.7287/peerj.preprints.3190v2>

Thanaki, J. (2018). *Machine Learning Solutions Expert techniques to tackle complex machine learning problems using Python*. Packt Publishing.

Vergara Cardona, T. (2022). *Modelo para la recomendación de fondos de inversión colectiva para personas naturales en el mercado financiero colombiano* [Tesis de Maestría, Universidad Nacional de Colombia Sede Medellín].

<https://repositorio.unal.edu.co/handle/unal/81029>

Zheng, A., y Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media, Inc.

Apéndices

Apéndice A

Características Conjuntos de Datos AUM FIC

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Subtipo	Catagórica	Subtipo del Fondo de Inversión Colectiva
cod_entidad	Catagórica	Código de la entidad fiduciaria asignado por la SFC
nombre_negocio	Catagórica	Nombre del Fondo de Inversión Colectiva
cod_negocio	Catagórica	Código del negocio registrado ante la SFC
AUM FIC	Numérica	Activos administrados por el Fondo de Inversión Colectiva
Inversionistas	Numérica	Número de inversionistas del Fondo de Inversión Colectiva

Apéndice B*Características Conjunto de Datos Agregados Monetarios Crediticios*

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Efectivo	Numérica	Corresponde al saldo de los billetes y monedas (de pesos colombianos) en poder del público. Además, en este agregado se incorporan los depósitos de agentes no financieros (distintos del Gobierno Nacional) en el Banco de la República
Reserva Bancaria	Numérica	Corresponde al saldo de billetes y monedas (de pesos colombianos) en la caja de las entidades depositarias, y sus depósitos en el Banco de la República en cuentas que sirven para cumplimiento del requisito de encaje
Total	Numérica	Suma del efectivo y la reserva bancaria
Del sector privado	Numérica	Saldo de cuentas corrientes del sector privado
Del sector público	Numérica	Saldo de cuentas corrientes del sector público
Total 1	Numérica	Suma de los saldos de cuentas corrientes del sector público y privado
Cuentas de ahorros	Numérica	Corresponde al saldo de los depósitos de este tipo que distintos agentes tienen en sociedades depositarias. Estos permiten al titular disponer en cualquier momento de los recursos para realizar pagos en comercios o retiros en efectivo en cajeros

		automáticos por medio de la tarjeta débito. La disponibilidad del dinero puede ser algo menor que las cuentas corrientes a cambio de una rentabilidad
CDT	Numérica	Corresponde al saldo de los Certificados de Depósito a Término ⁶ emitidos por las sociedades depositarias, excluye los CDT en poder del Banco de la República
TOTAL1	Numérica	Suma del saldo de cuentas de ahorros y CDT
Otros depósitos a la vista	Numérica	Corresponden al saldo de los fondos depositados en el sistema financiero que pueden ser retirados en cualquier momento, diferentes a cuentas corrientes y de ahorro
Depósitos fiduciarios	Numérica	Corresponden al saldo de los depósitos por servicios bancarios de recaudo
Cédulas	Numérica	Corresponde al saldo de las cédulas hipotecarias emitidas por el liquidado Banco Central hipotecario
Bonos	Numérica	Corresponde al saldo de los títulos de deuda emitidos por las SD y excluye los bonos en poder del Banco de la República
Repos con la DGCPTN	Numérica	Corresponde a las operaciones repo que realizan los establecimientos de crédito con la Dirección General de Crédito Público y del Tesoro Nacional
Repos con entidades no financieras y la DGCPTN	Numérica	Corresponde al saldo de los Repos, Simultáneas y Transferencia Temporal de Valores (TTV) con entidades no financieras

Total.2	Numérica	Suma de Repos con la DGCPTN y Repos con entidades no financieras y la DGCPTN
Total Depósitos en poder del público 1	Numérica	Suma de total de cuentas corrientes, cuentas de ahorro, CDT, otros depósitos a la vista, depósitos fiduciarios y bonos
M1	Numérica	Agregado monetario que incluye el Efectivo y los Depósitos en cuenta corriente
M2	Numérica	Agregado monetario que incluye M1, Cuasidineros y excluye los CDT y los bonos en poder del Banco de la República
M3	Numérica	Agregado monetario que incluye el Efectivo, los Depósitos en poder del público, los depósitos restringidos y excluye los CDT y los bonos en poder del Banco de la República
Depósitos restringidos	Numérica	Corresponde al saldo los depósitos en las entidades que entraron en liquidación durante la crisis de finales de la década del 90. A partir de febrero de 2005, se dejaron de considerar en la operación estadística puesto que su valor no era significativo. Se incluyen en el M3, pero no en el total de depósitos
CDT en poder del Banco de la República	Numérica	Corresponde a los CDTs adquiridos por el Banco de la República. Estos CDT no hacen parte de los agregados monetarios
Bonos en poder del Banco de la República	Numérica	Corresponde a los Bonos adquiridos por el Banco de la República. Estos Bonos no hacen parte de los agregados monetarios

Total Depósitos 2	Numérica	Suma de total de depósitos en poder de lo público, depósitos restringidos y CDT y Bonos en poder del banco de la república
-------------------	----------	--

Apéndice C

Características Conjunto de Datos Índice de Precios al Consumidor (IPC)

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Índice	Numérica	Índice de precios al consumidor
Inflación anual %	Numérica	Se mide tomando el IPC de un mes y calculando su variación frente al dato del mismo mes del año anterior
Inflación mensual %	Numérica	Se mide tomando el IPC de un mes y calculando su variación frente al dato del mes anterior
Inflación año corrido %	Numérica	Se mide tomando el IPC de un mes y calculando su variación frente al dato del cierre de año anterior

Apéndice D*Características Conjunto de Datos Tasa de Intervención de Política Monetaria*

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Tasa de intervención de política monetaria (%)	Numérica	Tasa a la cual el banco central interviene directamente en el mercado monetario para suministrar o extraer liquidez del sistema

Apéndice E. Características Conjunto de Datos Categorización FIC

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Fiduciaria	Categórica	Entidad fiduciaria que administra los activos de los FIC
Código_negocio	Categórica	Código del negocio registrado ante la SFC
Categoría	Categórica	Categoría asignada al FIC
Nombre_participacion	Categórica	Nombre del FIC

Apéndice F*Características Conjunto de Datos Tasas de Captación CDT*

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Tasa 30 días	Numérica	Tasa efectiva anual de CDT a 30 días
Monto A 30 días	Numérica	Monto en millones de pesos de CDT negociados a 30 días
Tasa Entre 31 y 44 días	Numérica	Tasa efectiva anual de CDT entre 31 y 44 días
Monto Entre 31 y 44 días	Numérica	Monto en millones de pesos de CDT negociados entre 31 y 44 días
Tasa A 45 días	Numérica	Tasa efectiva anual de CDT a 45 días
Monto A 45 días	Numérica	Monto en millones de pesos de CDT negociados a 45 días
Tasa Entre 46 y 59 días	Numérica	Tasa efectiva anual de CDT entre 46 y 59 días
Monto Entre 46 y 59 días	Numérica	Monto en millones de pesos de CDT negociados entre 46 y 59 días
Tasa A 60 días	Numérica	Tasa efectiva anual de CDT a 60 días
Monto A 60 días	Numérica	Monto en millones de pesos de CDT negociados a 60 días
Tasa Entre 61 y 89 días	Numérica	Tasa efectiva anual de CDT entre 61 y 89 días
Monto Entre 61 y 89 días	Numérica	Monto en millones de pesos de CDT negociados entre 61 y 89 días
Tasa A 90 días	Numérica	Tasa efectiva anual de CDT a 90 días
Monto A 90 días	Numérica	Monto en millones de pesos de CDT negociados a 90 días

Tasa Entre 91 y 119 días	Numérica	Tasa efectiva anual de CDT entre 91 y 119 días
Monto Entre 91 y 119 días	Numérica	Monto en millones de pesos de CDT negociados entre 91 y 119 días
Tasa A 120 días	Numérica	Tasa efectiva anual de CDT a 120 días
Monto A 120 días	Numérica	Monto en millones de pesos de CDT negociados entre a 120 días
Tasa Entre 121 y 179 días	Numérica	Tasa efectiva anual de CDT entre 121 y 179 días
Monto Entre 121 y 179 días	Numérica	Monto en millones de pesos de CDT negociados entre 121 y 179 días
Tasa A 180 días	Numérica	Tasa efectiva anual de CDT a 180 días
Monto A 180 días	Numérica	Monto en millones de pesos de CDT negociados a 180 días
Tasa Entre 181 y 359 días	Numérica	Tasa efectiva anual de CDT entre 181 y 359 días
Monto Entre 181 y 359 días	Numérica	Monto en millones de pesos de CDT negociados entre 181 y 359 días
Tasa A 360 días	Numérica	Tasa efectiva anual de CDT a 360 días
Monto A 360 días	Numérica	Monto en millones de pesos de CDT negociados a 360 días
Tasa Superiores a 360 días	Numérica	Tasa efectiva anual de CDT superiores a 360 días
Monto Superiores a 360 días	Numérica	Monto en millones de pesos de CDT negociados superiores a 360 días
Tasa Captaciones a través de CDT por red de oficinas	Numérica	Tasa de captación de CDT negociados a través de red de oficinas
Monto Captaciones a través de CDT por red de oficinas	Numérica	Monto de CDT negociados a través de red de oficinas

Tasa Captaciones a través de CDT por tesorería	Numérica	Tasa de captación de CDT negociados a través de tesorería
Monto Captaciones a través de CDT por tesorería	Numérica	Monto de CDT negociados a través de tesorería

Apéndice G

Características Conjunto de Datos Valor COLTES

Variable	Tipo de Variable	Definición
Fecha	Datetime	Fecha de recolección y/o publicación de la información
Valor hoy	Numérica	Valor del Coltes
Valor ayer	Numérica	Valor del Coltes día anterior
Variación absoluta	Numérica	Valoración absoluta entre días Coltes
Variación porcentual	Numérica	Valoración porcentual entre días Coltes
Variación 12 meses	Numérica	Variación del último año Coltes
Variación año	Numérica	Variación año corrido Coltes
Tir	Numérica	Tasa Interna de Retorno
Duración	Numérica	Duración

Apéndice H*Características Conjunto de Datos Valor Dólar*

Variable	Tipo de Variable	Definición
Valor	Numérica	Valor de la TRM del día
Vigencia desde	Numérica	Fecha desde la cual tiene vigencia la TRM
Vigencia hasta	Numérica	Fecha hasta la cual tiene vigencia la TRM

Apéndice I

Fuentes de Datos

Las fuentes de datos usados para el desarrollo de este estudio se encuentran dentro de un repositorio de GitHub, al cual se podrá acceder a través del siguiente enlace:

<https://github.com/andress2800/Prediccion-AUM-FIC/tree/main/Fuentes%20de%20datos>

Apéndice J

Jupyter Notebook Análisis Exploratorio De Datos

El Jupyter notebook correspondiente al análisis exploratorio de datos se encuentran en un repositorio de GitHub al cual se puede acceder a través del siguiente enlace

<https://github.com/andress2800/Prediccion-AUM-FIC/tree/main/Analisis%20Exploratorio>

Apéndice K

Jupyter Notebook Análisis Exploratorio De Datos

Los Jupyter notebooks en los que se entrenaron y probaron los modelos se encuentran en un repositorio de GitHub al cual se puede acceder a través del siguiente enlace

<https://github.com/andress2800/Prediccion-AUM-FIC/tree/main/Modelado%20de%20Datos>