



**Motor prescriptivo de pricing de CDTs basado en curvas de mercado y aprendizaje automático:
caso Corficolombiana.**

Anggie Tatiana Moreno Abella

Carlos Andrés Aranda Barreto

Universidad EAN

Facultad de Ingeniería

Maestría en Ciencia de Datos

Bogotá, Colombia

06 de junio de 2026

**Motor prescriptivo de pricing de CDTs basado en curvas de mercado y aprendizaje automático:
caso Corficolombiana.**

Anggie Tatiana Moreno Abella

Carlos Andrés Aranda Barreto

Trabajo de grado presentado como requisito para optar al título de:

Magíster en Ciencia de Datos

Director (a):

Leonardo Pérez Cortes

Modalidad:

Trabajo Dirigido

Universidad EAN

Facultad de Ingeniería

Maestría en Ciencia de Datos

Bogotá, Colombia

06 de junio de 2026

Nota de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Bogotá, 06/junio/2026

Anggie Tatiana Moreno Abella

Al universo, por conspirar siempre a mi favor y ponerme en el lugar exacto donde tenía que estar.

A mi papi, quien partió antes de ver este logro, pero cuya presencia sigo sintiendo en cada paso que doy. Este título también es tuyo, y espero que, desde donde estés, te sientas orgulloso de mí siempre.

A mi mami, por su amor incondicional y por enseñarme que la constancia es la forma más silenciosa de la valentía.

A mi hermano, por estar siempre.

A mi perrita y gatico, por las noches de estudio acompañadas de sus cabezas apoyadas en mis piernas y por recordarme, en medio del cansancio, la importancia de detenerme y recargar el alma.

Y a Carlos, por su apoyo, su paciencia y por estar presente a lo largo de este camino, brindándome ánimo y confianza en cada etapa de este logro

Carlos Andrés Aranda Barreto

Al universo, por haberme dado exactamente las personas que necesitaba en el momento justo.

A mi madre hermosa, que partió físicamente el día que cumplí 18 años, pero que jamás se ha ido de mi corazón. Hoy, cada logro lleva su nombre y su sonrisa. Esta tesis es también para ella, porque desde donde esté, sigue siendo mi primera y más inspiradora fuerza.

A mis primas, sus esposos e hijos, a mi padrino y a mi tía, que son mi familia, mi raíz y mi red.

A mis amigos de toda la vida, Hernán, Danilo, Juan y Jhon; gracias a toda mi familia y amigos por creer en mí incluso cuando yo dudaba.

Y a Anggie, por su apoyo técnico, sus aportes intelectuales y su disposición constante durante el desarrollo de esta tesis.

Agradecimientos

La culminación de este trabajo no habría sido posible sin el concurso generoso de muchas personas e instituciones que, de distintas maneras, hicieron posible este camino.

A la Universidad EAN y a su planta de profesores de la Maestría en Ciencia de Datos, por la formación rigurosa, los retos académicos y la orientación permanente. Cada clase, cada retroalimentación y cada exigencia dejó una huella que se refleja en estas páginas.

A nuestro director Leonardo Pérez Cortes, por su guía, su disposición y por ayudarnos a encontrar la dirección correcta cuando el camino se veía difuso.

A Corficolombiana, por abrir sus puertas a esta investigación y por permitir que un problema real de negocio se convirtiera en el corazón de este trabajo. Llevar años dentro de esta organización nos dio no solo el contexto, sino también la convicción de que la propuesta tenía sentido y valor.

A nuestro jefe Humberto Cerpa, por la confianza depositada al respaldar este proyecto dentro de la empresa, por los espacios cedidos y por entender que el desarrollo profesional y el académico no son caminos separados.

Y finalmente, nos agradecemos el uno al otro. Construir una tesis de maestría es un proceso que pone a prueba la paciencia, la comunicación y el compromiso. Hacerlo juntos —como compañeros de trabajo y de estudio— fue a la vez el mayor reto y privilegio de este proceso.

Resumen

Esta investigación presenta una propuesta de intervención empresarial para fortalecer el pricing de los Certificados de Depósito a Término (CDTs) en Corficolombiana. A partir del análisis de las operaciones de la entidad, surge la siguiente pregunta de investigación: ¿cómo puede Corficolombiana tomar mejores decisiones sobre las tasas de sus CDTs, de manera más consistente, más trazable y menos dependiente del criterio de unas pocas personas? Hoy, ese proceso depende en buena medida del juicio experto individual, carece de una referencia analítica integrada y no siempre deja registro claro de por qué se ofreció una tasa determinada, todo esto en un contexto donde el costo del fondeo ha afectado materialmente los resultados de tesorería.

Para responder a esa pregunta, se desarrolla un motor prescriptivo de pricing soportado en ciencia de datos y aprendizaje automático. El motor integra información de curvas de mercado, series macroeconómicas y señales competitivas; construye referencias comparables por plazo e indexación; y calcula una medida relativa de posicionamiento conocida como basis de emisión, que permite comparar cada captación frente a lo que el mercado ofrece en condiciones equivalentes. Sobre esa base, se desarrollaron modelos supervisados que recomiendan tasas y spreads por plazo e indexación, con resultados técnicos suficientes para apoyar la decisión en condiciones normales de mercado.

El diagnóstico organizacional mostró que la principal brecha de Corficolombiana no es la falta de información, sino la ausencia de una capa analítica integrada, reproducible y explicable que convierta señales dispersas de mercado en decisiones consistentes. En respuesta a ello, el trabajo formula un plan de intervención orientado a institucionalizar la solución mediante gobierno de datos, validación del modelo, adopción por el usuario de negocio y monitoreo continuo. Se concluye que la propuesta aporta valor no solo por su desempeño técnico, sino por su capacidad de fortalecer la inteligencia decisonal sobre el fondeo bajo criterios de uso responsable, supervisión humana y trazabilidad.

Palabras clave: pricing de CDTs, costo de fondeo, tesorería, curvas de mercado, aprendizaje automático, intervención empresarial.

Abstract

This study presents a business intervention proposal aimed at strengthening certificate of deposit (CDT) pricing in Corficolombiana. Based on an analysis of the institution's operations, the following research question emerges: how can Corficolombiana make better decisions about its CDT rates, in a more consistent, traceable, and less expert-dependent way? Today, that process relies heavily on individual expert judgment, lacks an integrated analytical reference, and does not always leave a clear record of why a specific rate was offered — all of this in a context where funding costs have materially affected treasury results.

To answer that question, this study develops a prescriptive pricing engine supported by data science and machine learning. The engine integrates market curve data, macroeconomic series, and competitive signals; builds comparable references by maturity and indexation; and computes a relative pricing measure known as issuance basis, which allows each deposit to be benchmarked against what the market offers under equivalent conditions. On that basis, supervised models were developed to recommend rates and spreads by maturity and indexation, with technical results sufficient to support decision-making under normal market conditions.

The organizational diagnosis revealed that the institution's main gap is not the lack of data, but rather the absence of an integrated, reproducible, and explainable analytical layer capable of turning dispersed market signals into consistent pricing decisions. In response, the study proposes an implementation plan focused on data governance, model validation, business-user adoption, and continuous monitoring. The study concludes that the proposal adds value not only because of its

technical performance, but also because of its potential to strengthen funding-related decision intelligence under principles of responsible use, human oversight, and traceability.

Keywords: CDT pricing, funding cost, treasury, market curves, machine learning, business intervention.

Tabla de Contenido

Tabla de Tablas.....	16
Tabla de Figuras.....	18
Introducción	20
Objetivos	22
<i>Objetivo General</i>	<i>22</i>
<i>Objetivos Específicos</i>	<i>22</i>
Justificación	23
Marco Institucional.....	25
<i>Presentación de la Empresa</i>	<i>25</i>
<i>Reseña histórica</i>	<i>26</i>
<i>Referentes Estratégicos.....</i>	<i>26</i>
<i>Estructura Organizacional</i>	<i>27</i>
<i>Productos y Servicios.....</i>	<i>29</i>
Posicionamiento en el Mercado	30
<i>Ubicación del proyecto dentro de la operación de Corficolombiana</i>	<i>31</i>
Marco de Referencia.....	32
<i>Certificados de Depósito a Término (CDTs): instrumento, modalidades e indexadores</i>	<i>32</i>
<i>Precia y los datos de mercado en Colombia</i>	<i>33</i>

<i>Curvas de mercado, estructura temporal de tasas y basis de emisión</i>	<i>34</i>
Estructura temporal de tasas.....	34
Curva swap y referencia monetaria	35
Basis de emisión como medida relativa de pricing.....	36
<i>Ciencia de datos aplicada al pricing financiero: pipelines, ingeniería de datos y modelado.....</i>	<i>39</i>
El pipeline de datos como artefacto analítico	39
Ingeniería de características financieras (Feature Engineering).....	40
Algoritmos de aprendizaje automático supervisado para regresión.....	41
Evaluación y selección de modelos.....	43
Interpretabilidad de modelos de aprendizaje automático	44
<i>Gobierno de datos en el sector financiero</i>	<i>45</i>
Definición y principios.....	45
Linaje de datos y trazabilidad	46
Calidad del dato	47
<i>Riesgo de modelo y gobierno del modelo en entidades financieras</i>	<i>47</i>
<i>Marco normativo colombiano relevante para el pricing de CDTs.....</i>	<i>49</i>
Diseño Metodológico.....	51
<i>Enfoque del estudio y caracterización metodológica</i>	<i>51</i>
<i>Fases para alcanzar los objetivos del estudio</i>	<i>52</i>
<i>Metodología de diagnóstico empresarial seleccionada</i>	<i>53</i>
<i>Población, muestra, instrumentos y validación</i>	<i>55</i>
<i>Proceso de validación del instrumento</i>	<i>56</i>

<i>Fuentes de información, procedimientos y técnicas para recoger y analizar la información....</i>	<i>58</i>
<i>Ficha técnica de la intervención.....</i>	<i>60</i>
<i>Variables de estudio, evaluación e interpretabilidad.....</i>	<i>61</i>
<i>Resultados del instrumento aplicado a informantes clave</i>	<i>62</i>
Hallazgos cualitativos sobre el proceso actual	63
Hallazgos cuantitativos: frecuencias del proceso actual	64
Uso de fuentes de información.....	65
Criterios de adopción de la herramienta	65
<i>Metodología de analítica y ciencia de datos: CRISP-DM</i>	<i>66</i>
<i>Metodología para formular el plan de mejora de la intervención.....</i>	<i>67</i>
<i>Gobierno de datos, ética y gobierno del modelo</i>	<i>67</i>
<i>Resumen de la relación objetivos–metodología</i>	<i>68</i>
Diagnóstico Organizacional.....	70
<i>Propósito y alcance del diagnóstico.....</i>	<i>70</i>
<i>Diagnóstico externo</i>	<i>71</i>
Tendencias del Sector y Entorno Macroeconómico	71
Competencia, información y formación de tasas	74
Diagnóstico interno del proceso actual de pricing de CDTs	75
Situación AS-IS del proceso.....	75
Capacidades instaladas	76
Brechas identificadas	77
Análisis de las Cinco Fuerzas de Porter	78

Amenaza de Nuevos Entrantes.	78
Poder de Negociación de los Proveedores.	79
Poder de Negociación de los Clientes.	79
Amenaza de Productos Sustitutos.	79
Rivalidad Entre Competidores.	80
Análisis PESTEL.....	81
Político.	81
Económico.....	81
Social.	81
Tecnológico.	82
Ambiental.....	82
Legal.	82
Análisis DOFA (Fortalezas, Oportunidades, Debilidades y Amenazas)	83
Fortalezas.	83
Oportunidades.	84
Debilidades.....	84
Amenazas.....	84
Priorización de problemas	85
Conclusión del diagnóstico	86
Desarrollo del pipeline analítico y resultados del modelado	88
<i>Arquitectura general del pipeline.....</i>	<i>88</i>
Capa 1. Ingesta de información macroeconómica y de mercado.....	90
Capa 2. Construcción de maestros de TES y CDTs	91

Capa 3. Cálculo de la prima o basis de emisión	93
Capa 4. Construcción de curvas soberanas interpoladas mediante PCHIP	95
Capa 5. Procesamiento de curvas cero cupón, curvas swap y betas	97
Capa 6. Feature engineering macrofinanciero y construcción del dataset maestro	99
Capa 7. Preprocesamiento final para machine learning y control de leakage	102
Capa 8. Modelado temporal supervisado y selección de modelos	105
Resultados del modelado temporal.....	106
Interpretación metodológica de los resultados.....	108
Capa 9. Interpretabilidad, diagnóstico de validación y análisis de drivers del modelo	108
Métodos de interpretabilidad implementados	109
Resultados de interpretabilidad por escenario	110
Tasa Fija.....	110
IBR	112
IPC	114
Basis	116
Capa 10. Despliegue funcional mediante aplicación Streamlit.....	119
Pestaña Datos	119
Pestaña Torneo	120
Pestaña Interpretabilidad	123
Pestaña Motor de Pricing.....	125
Escalabilidad, arquitectura objetivo y operación MLOps	127
<i>Validación funcional del prototipo con usuarios de mesa</i>	<i>128</i>
Resultados cualitativos de la validación funcional	130

Plan de Intervención	133
<i>Objetivo del plan de implementación</i>	<i>133</i>
<i>Principios de implementación</i>	<i>133</i>
<i>Articulación entre diagnóstico, pipeline y plan de intervención</i>	<i>134</i>
Fase 1. Formalización funcional y de gobierno.....	134
Fase 2. Industrialización del pipeline de datos	135
Fase 3. Validación analítica y selección del modelo operativo.....	135
Fase 4. Despliegue y adopción por usuario de negocio.....	136
Fase 5. Monitoreo y mejora continua.....	136
<i>Estructura del plan por componentes</i>	<i>137</i>
<i>Relación entre brechas priorizadas, acciones y entregables.....</i>	<i>139</i>
<i>Indicadores de éxito</i>	<i>141</i>
<i>Resultado esperado de la intervención</i>	<i>142</i>
<i>Cronograma propuesto de implementación</i>	<i>142</i>
<i>Responsables propuestos.....</i>	<i>143</i>
<i>Recursos mínimos requeridos</i>	<i>144</i>
<i>Riesgos de implementación y medidas de mitigación</i>	<i>145</i>
Conclusiones	147
Recomendaciones.....	150
Referencias.....	152

Anexos	160
<i>Anexo A – Carta de Aval de la Empresa Para Realizar la Intervención</i>	<i>160</i>
<i>Anexo B - Instrumento de medición final aplicado a informantes clave.....</i>	<i>161</i>
Consentimiento informado.....	161
A. Información general y contexto de participación.....	162
B. Experiencias recientes del proceso actual	162
C. Estandarización, trazabilidad y dependencia de criterio experto	162
D. Uso real de información y necesidades de datos	163
E. Criterios de adopción del motor prescriptivo	164
F. Cierre	165
<i>Anexo C - Matriz de correlación de cada escenario.</i>	<i>166</i>
<i>Anexo D - Glosario técnico y diccionario de variables del pipeline</i>	<i>168</i>

Tabla de Tablas

Tabla 1 Ficha de Caracterización.....	52
Tabla 2 Fases del diseño metodológico y relación con los objetivos.....	53
Tabla 3 Herramientas metodológicas para el diagnóstico externo e interno	54
Tabla 4 Población y muestra	56
Tabla 5 Instrumentos del estudio y criterio de validación.....	58
Tabla 6 Fuentes de información y trazabilidad del pipeline	59
Tabla 7 Ficha técnica de la intervención	60
Tabla 8 Variables de estudio	61
Tabla 9 Frecuencias del proceso actual de pricing según informantes clave	64
Tabla 10 Aplicación de CRISP-DM al pipeline analítico del proyecto	66
Tabla 11 Relación Objetivos–Metodología	68
Tabla 12 Síntesis del Diagnóstico Externo	74
Tabla 13 Levantamiento AS-IS del Proceso Actual de Pricing de CDTs.....	76
Tabla 14 Brechas Diagnósticas del Proceso Actual.....	78
Tabla 15 Resumen Análisis de las Cinco Fuerzas de Porter	80
Tabla 16 Resumen Análisis PESTEL	82
Tabla 17 Resumen Análisis DOFA	85
Tabla 18 Priorización de Problemas del Diagnóstico	86
Tabla 19 Arquitectura del Pipeline Analítico por Componente	88
Tabla 20 Variables Feature Engineering	100
Tabla 21 Resultados del Modelado.....	107
Tabla 22 Resumen Interpretabilidad	118
Tabla 23 Arquitectura Objetivo Para Escalabilidad y Operación MLOps.....	128

Tabla 24 Instrumento de Validación Funcional del Prototipo Con Usuarios de la Mesa	129
Tabla 25 Síntesis de Hallazgos de la Validación Funcional del Prototipo	131
Tabla 26 Pipeline Analítico Integral	137
Tabla 27 Trazabilidad del Dato Hasta la Recomendación	138
Tabla 28 Gobierno y Control del Motor Prescriptivo.....	138
Tabla 29 Mapeo Entre Brechas Priorizadas, Componentes Técnicos y Entregables de Intervención	139
Tabla 30 Cronograma Propuesto de Implementación del Motor Prescriptivo de Pricing de CDTs	142
Tabla 31 Responsables Propuestos Para la Implementación	143
Tabla 32 Recursos Mínimos Requeridos Para la Implementación.....	144
Tabla 33 Riesgos de Implementación y Mitigación	145
Tabla 34 Glosario Técnico y Diccionario de Variables del Pipeline.....	168

Tabla de Figuras

Figura 1 Esquema Simplificado de la Estructura Organizacional de Corficolombiana	28
Figura 2 Índice de Precios al Consumidor	71
Figura 3 Crecimiento PIB Real, Trimestral Años 2023, 2024 y 2025	72
Figura 4 Resultados Pipeline 2.2_Maestro_CDT SX.py	93
Figura 5 Resultados Pipeline 3.1_Motor_Basis.py	95
Figura 6 Resultados Pipeline 3.2_Curva_TES_Pchip.py	97
Figura 7 Resultados Pipeline 4.1_Maestro_CceroC.py	98
Figura 8 Resultados Pipeline 4.2_Feature_Engineering_Macro.py	101
Figura 9 Resultados Pipeline 5_Preprocesamiento_ML.py	104
Figura 10 Importancia por Permutación Para el Escenario de Tasa Fija.....	110
Figura 11 Diagnóstico de Validación Para el Escenario de Tasa Fija.....	111
Figura 12 Importancia por Permutación Para el Escenario de IBR	113
Figura 13 Diagnóstico de Validación Para el Escenario de IBR	114
Figura 14 Importancia por Permutación Para el Escenario de IPC	115
Figura 15 Diagnóstico de Validación Para el Escenario de IPC	116
Figura 16 Importancia por Permutación Para el Escenario de Basis	117
Figura 17 Diagnóstico de Validación Para el Escenario de Basis	118
Figura 18 Aplicación Pestaña Datos	120
Figura 19 Aplicación Pestaña Torneo – Escenario Tasa Fija.....	121
Figura 20 Aplicación Pestaña Torneo – Escenario IBR	121
Figura 21 Aplicación Pestaña Torneo – Escenario IPC	122
Figura 22 Aplicación Pestaña Torneo – Escenario Basis	122
Figura 23 Aplicación Pestaña Interpretabilidad – Escenario Tasa Fija.....	123

Figura 25 Aplicación Pestaña Interpretabilidad – Escenario IPC	124
Figura 24 Aplicación Pestaña Interpretabilidad – Escenario IBR	124
Figura 26 Aplicación Pestaña Interpretabilidad – Escenario Basis	125
Figura 27 Aplicación Pestaña Motor de Pricing – Cálculo de Precio Según Parámetros	126
Figura 28 Aplicación Pestaña Motor de Pricing – Cálculo Curva Completa	126
Figura 29 Matriz de Correlación Para Tasa Fija	166
Figura 30 Matriz de Correlación Para IBR	166
Figura 31 Matriz de Correlación Para IPC	167
Figura 32 Matriz de Correlación Para Basis	167

Introducción

En diciembre de 2024, la tesorería de Corficolombiana registró una tasa efectiva promedio de costo en CDTs de 11,01 % y cerró el año con pérdidas por COP 56.453 millones (Corficolombiana, 2025b). Detrás de esas cifras hay una pregunta que este trabajo intenta responder: ¿existe una forma más disciplinada, trazable y basada en datos de definir las tasas de captación?

La definición de tasas de captación es una decisión clave para la tesorería de cualquier entidad financiera, porque afecta de manera directa su competitividad comercial, el costo de fondeo y la estabilidad del margen financiero. En Colombia, este reto ha cobrado mayor relevancia en los últimos años, en un entorno marcado por desaceleración económica, descenso gradual de la inflación y alta sensibilidad de los mercados a los cambios en tasas de interés. De acuerdo con el DANE, en 2024 el crecimiento provisional del PIB fue de 1,5 % (DANE, 2025b), mientras que la inflación anual a diciembre se ubicó en 5,20 % (DANE, 2025a). En un contexto así, definir tasas de captación no puede depender únicamente de criterio experto o de referencias parciales del mercado, sino de herramientas que permitan tomar decisiones más consistentes, comparables y trazables.

Bajo esta premisa, el presente trabajo dirigido propone una intervención empresarial orientada a fortalecer el pricing de los Certificados de Depósito a Término (CDTs) en Corficolombiana mediante un motor prescriptivo apoyado en ciencia de datos, curvas de mercado y aprendizaje automático. La elección del caso no es arbitraria: además de las pérdidas de tesorería ya mencionadas, el reporte del primer trimestre de 2025 mostró que el margen neto de tesorería continuó siendo negativo, lo que confirma que el problema sigue vigente y requiere una respuesta más estructurada (Corficolombiana, 2025d).

La relevancia del proyecto también se entiende a la luz del entorno sectorial. El Banco de la República ha advertido que, si bien el sistema financiero colombiano mantiene niveles adecuados de

capital y liquidez, la mayor exposición a deuda pública aumenta la vulnerabilidad de las entidades frente a movimientos adversos en la prima de riesgo soberano y en los mercados financieros locales (Banco de la República, 2024b). Para una entidad como Corficolombiana, con participación en renta fija, divisas, derivados y fondeo, esto implica que la definición de tasas de captación debe apoyarse en una lectura más integrada de las curvas de mercado, de las señales macroeconómicas y de la presión competitiva.

Desde el punto de vista metodológico, la propuesta se apoya en un pipeline reproducible que automatiza la adquisición, transformación e integración de insumos de mercado y variables macroeconómicas, construye referencias comparables por plazo e indexación, estima medidas relativas de pricing como el basis de emisión y entrena modelos capaces de apoyar la recomendación de tasas y spreads. Sin embargo, el valor del trabajo no está solo en la predicción, sino en la posibilidad de convertir esa capacidad analítica en una herramienta útil, explicable y trazable para el negocio. Por eso, la intervención incorpora lineamientos de gobierno de datos, validación y uso responsable de modelos, en coherencia con referentes como SR 11-7 y NIST AI RMF 1.0 (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011; Tabassi, 2023).

En ese marco, el documento se organiza en capítulos. Primero, presenta el contexto institucional y sectorial del caso de estudio. Segundo, desarrolla el marco de referencia que sustenta la relación entre fondeo, captación a plazo, curvas de mercado y analítica avanzada. Tercero, expone el diseño metodológico de la intervención. Cuarto, desarrolla el diagnóstico organizacional del proceso actual de pricing de CDTs. Finalmente, formula un plan de intervención orientado a institucionalizar el motor prescriptivo dentro de la dinámica de tesorería, y cierra con las conclusiones y recomendaciones del estudio.

Objetivos

Objetivo General

Diseñar una propuesta de intervención para el pricing prescriptivo de CDTs en Corficolombiana, soportada en ciencia de datos, curvas de mercado y variables macroeconómicas y competitivas, con el fin de mejorar la consistencia, trazabilidad y eficiencia de la decisión comercial asociada al costo de fondeo.

Objetivos Específicos

Revisar los referentes teóricos, sectoriales y regulatorios relacionados con el pricing de captaciones, la gestión del fondeo, las curvas de mercado y el uso de analítica avanzada en entidades financieras, con el fin de identificar las variables y criterios que deben sustentar la propuesta.

Diagnosticar el proceso actual de definición de tasas y spreads de CDTs en Corficolombiana, así como sus principales brechas frente a las mejores prácticas analíticas y de mercado, para establecer el punto de partida de la intervención.

Estructurar un pipeline reproducible de datos que integre información de mercado, variables macroeconómicas y señales competitivas, con el fin de soportar la construcción de un motor analítico para la estimación y recomendación de tasas o spreads de CDTs.

Desarrollar y validar un motor prescriptivo de pricing de CDTs, basado en técnicas de aprendizaje automático e interpretabilidad, que permita generar recomendaciones consistentes y trazables para distintos escenarios de plazo e indexación.

Formular lineamientos para la adopción del motor prescriptivo en Corficolombiana y evaluar su utilidad práctica mediante una prueba con usuario de negocio, con el fin de identificar fortalezas, limitaciones y recomendaciones de mejora para su desarrollo futuro.

Justificación

La investigación responde a una necesidad concreta de Corficolombiana: contar con una forma más disciplinada, trazable y basada en datos para definir las tasas de sus CDTs, en un contexto en el que el costo del fondeo ha afectado de manera directa los resultados financieros y la toma de decisiones de tesorería. De acuerdo con el informe de gestión de 2024, la tesorería de la entidad “cerró 2024 con pérdidas por valor de COP 56.453 millones” (Corficolombiana, 2025b, p. 26), asociadas principalmente al carry negativo derivado del alto costo del financiamiento; adicionalmente, el mismo informe reporta una “tasa efectiva promedio de costo en Certificados de Depósito a Término (CDT) de 11,01 %” (Corficolombiana, 2025b, p. 26), lo que refuerza la relevancia de fortalecer el pricing de captaciones mediante herramientas analíticas más consistentes.

La pertinencia del proyecto también se relaciona con la escala y complejidad del negocio de tesorería y mercado de la organización. En 2024, Corficolombiana mantuvo una participación relevante en el mercado local de deuda pública al ubicarse en el puesto 10 dentro del esquema de Creadores de Mercado y alcanzó una participación de 16,95 % en el mercado spot peso-dólar (Corficolombiana, 2025b, p. 26). Estas cifras muestran que mejoras en la definición de tasas y spreads pueden tener un impacto material sobre la competitividad del fondeo y la eficiencia comercial (Corficolombiana, 2025b). Además, aunque en el primer trimestre de 2025 el margen neto de tesorería mostró una mejora frente al mismo periodo de 2024, permaneció en terreno negativo, lo que sugiere que el problema sigue vigente y requiere una respuesta más estructurada (Corficolombiana, 2025d).

Desde una perspectiva sectorial, el Banco de la República ha advertido que, aunque los establecimientos de crédito en Colombia mantienen niveles adecuados de capital y liquidez, una mayor exposición a títulos de deuda pública incrementa su vulnerabilidad frente a movimientos adversos en la prima de riesgo soberano y en los mercados financieros locales (Banco de la República, 2024a). En ese entorno, un motor prescriptivo de pricing basado en curvas de mercado, señales macroeconómicas y

variables competitivas puede aportar mayor consistencia, oportunidad y comparabilidad a la toma de decisiones de captación.

Finalmente, la justificación metodológica del proyecto se apoya en la literatura institucional sobre inteligencia artificial y gobierno de modelos en el sector financiero. La Financial Stability Board ha señalado que la adopción de inteligencia artificial ofrece oportunidades de eficiencia y creación de valor, pero también introduce riesgos asociados con la gobernanza, la privacidad, la ética y el riesgo de modelo (FSB, 2024). De forma complementaria, la guía SR 11-7 subraya la necesidad de validación y control formal de los modelos cuantitativos (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011), mientras que el marco NIST AI RMF promueve el desarrollo y uso confiable y responsable de sistemas de inteligencia artificial (Tabassi, 2023). En consecuencia, el aporte del proyecto no se limita a recomendar tasas, sino que propone una solución analítica con criterios de trazabilidad, interpretabilidad, supervisión humana y lineamientos de gobierno de datos y modelo.

Marco Institucional

Presentación de la Empresa

La Corporación Financiera Colombiana S.A. (Corficolombiana) es un establecimiento de crédito de carácter privado, con domicilio principal en Bogotá y sometido a la vigilancia de la Superintendencia Financiera de Colombia. La entidad hace parte del conglomerado financiero Aval y del Grupo Empresarial Sarmiento Angulo, y su objeto social comprende la realización de los actos y contratos autorizados a las corporaciones financieras por el Estatuto Orgánico del Sistema Financiero (Corficolombiana, 2025b). De acuerdo con sus estados financieros separados y con su informe de gestión, Corficolombiana combina un rol de holding de inversiones con actividades especializadas en tesorería, banca de inversión, servicios financieros y análisis de mercado (Corficolombiana, 2025a).

Desde el punto de vista del problema de investigación, la organización resulta especialmente pertinente porque la tesorería y la gestión del fondeo ocupan un lugar material dentro de su desempeño financiero. El informe CorfiSostenible 2024 muestra que la entidad mantiene posiciones relevantes en deuda pública, derivados, divisas y captaciones, y que el costo de fondeo vía CDT sigue siendo una variable crítica para el margen financiero (Corficolombiana, 2025b). Por ello, la definición de tasas y spreads de captación no es una decisión marginal, sino un proceso que incide de manera directa en la competitividad comercial, la rentabilidad y la gestión integral del balance.

Para efectos de este trabajo, la unidad de análisis organizacional se concentra en la Vicepresidencia de Tesorería y en los procesos asociados a captación, mesas de dinero, ALM y gestión de fondeo. Esta delimitación es consistente con el alcance de la propuesta, dado que el proyecto no busca intervenir la totalidad del modelo corporativo de Corficolombiana, sino fortalecer la toma de decisiones sobre pricing de CDTs mediante una infraestructura analítica reproducible y trazable.

Reseña histórica

Corficolombiana fue constituida el 27 de noviembre de 1961 mediante escritura pública No. 5710 de la Notaría Primera del Círculo de Cali. Su trayectoria institucional refleja la evolución del sistema financiero colombiano y la transformación de las corporaciones financieras hacia modelos más complejos de intermediación, inversión y estructuración financiera. A lo largo de su historia, la organización ha ampliado su alcance mediante fusiones, reorganizaciones societarias y fortalecimiento de líneas de negocio especializadas (Corficolombiana, 2025a).

Durante las décadas siguientes, la entidad consolidó su crecimiento por medio de adquisiciones y procesos de integración que fortalecieron su posición en el sector financiero colombiano. Un hito decisivo se produjo en diciembre de 2005, cuando se formalizó la fusión entre la Corporación Financiera del Valle S.A. y la Corporación Financiera Colombiana S.A.; en esa misma operación, la entidad absorbente adoptó la razón social Corporación Financiera Colombiana S.A. y trasladó su domicilio principal a Bogotá. Posteriormente, nuevas integraciones contribuyeron a consolidar su perfil corporativo y de inversión (Corficolombiana, s/f-b).

En las dos últimas décadas, la Corporación ha fortalecido su presencia en sectores estratégicos como infraestructura, energía y gas, agroindustria, hotelería y servicios financieros, al tiempo que ha consolidado negocios de tesorería y banca de inversión (Corficolombiana, s/f-b). Esta trayectoria histórica es relevante para la presente investigación porque explica por qué la organización combina hoy capacidades de inversión estratégica con capacidades financieras avanzadas, entre ellas las requeridas para desarrollar analítica aplicada a fondeo, tasas y spreads.

Referentes Estratégicos

La misión de Corficolombiana es “trabajamos e invertimos en el progreso del país”, reflejando su compromiso con el desarrollo económico y social de Colombia. Su visión es ser reconocida “como una

empresa líder en sostenibilidad que crea valor económico, social y ambiental en los sectores en los que invierte, generando bienestar a la comunidad y a la región” (Corficolombiana, s/f-d, sección “Nuestra Filosofía”). Para lograrlo, la corporación adopta como principios rectores la transparencia, la sostenibilidad, la innovación, la excelencia, el respeto y el trabajo colaborativo (Corficolombiana, s/f-c).

El marco estratégico de Corficolombiana se encuentra organizado alrededor de la estrategia CorfiSostenible, la cual articula generación de valor, sostenibilidad, gobierno corporativo y gestión del riesgo. El informe de gestión de 2024 y la política de sostenibilidad de la entidad muestran que la organización concibe la sostenibilidad como un principio transversal para la toma de decisiones, la administración de riesgos y el relacionamiento con sus grupos de interés. En este sentido, la rentabilidad no se plantea como un objetivo aislado, sino como parte de una creación de valor que debe ser consistente con criterios de transparencia, responsabilidad y visión de largo plazo (Corficolombiana, 2025b).

Este marco estratégico es particularmente importante para el proyecto de investigación, porque el motor prescriptivo propuesto no se limita a estimar una tasa, sino que busca fortalecer un proceso de decisión comercial dentro de una organización regulada y altamente expuesta al riesgo de mercado. En consecuencia, la solución debe alinearse con principios de buen gobierno, trazabilidad, consistencia metodológica y supervisión humana. La pertinencia del proyecto, entonces, no solo se deriva de una necesidad de eficiencia financiera, sino también de la obligación institucional de tomar decisiones basadas en datos de forma responsable y auditable (Banco de la República, 2024a).

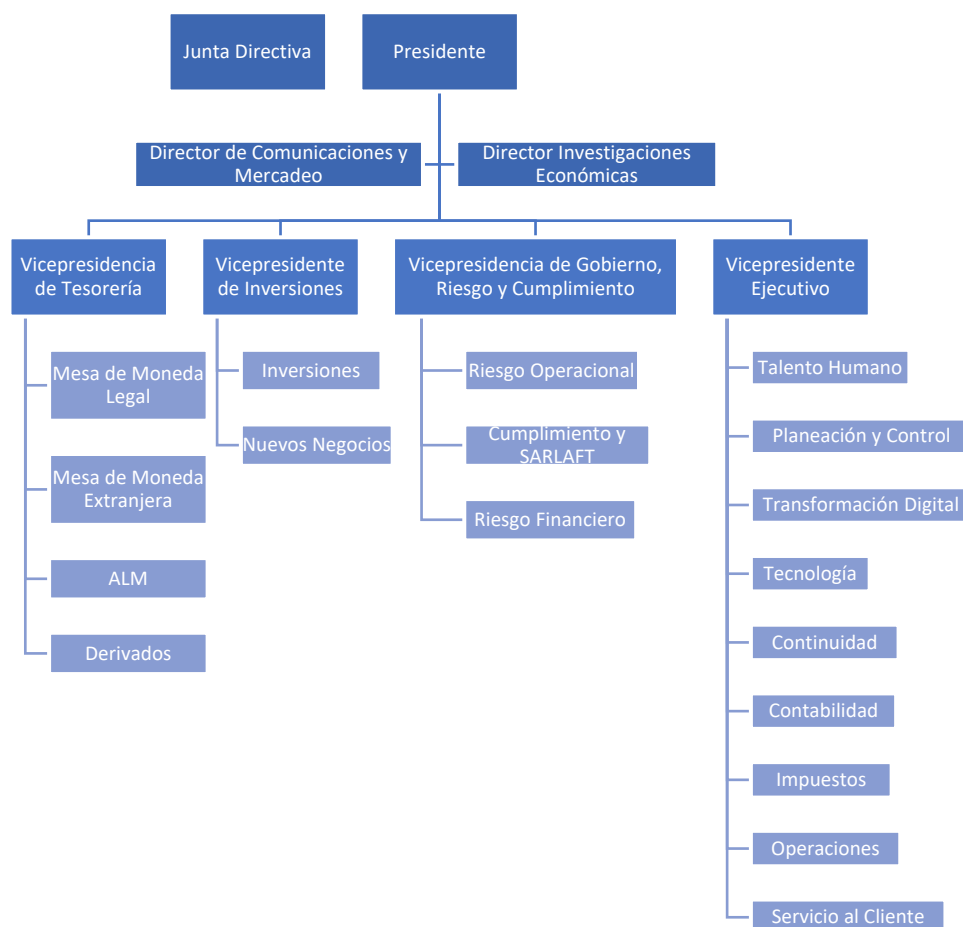
Estructura Organizacional

Corficolombiana se organiza bajo una estructura funcional con mecanismos de coordinación transversal entre vicepresidencias y áreas de apoyo. En la cúspide se encuentra la Junta Directiva, seguida por la Presidencia y por un conjunto de vicepresidencias corporativas que articulan el gobierno,

la operación y el control. El organigrama general publicado en febrero de 2025 (Figura 1) muestra, entre otras, la Vicepresidencia de Tesorería, la Vicepresidencia de Inversiones, la Vicepresidencia Corporativa de Gobierno, Riesgo y Cumplimiento y la Vicepresidencia Ejecutiva, apoyadas por áreas como Auditoría, Secretaría General, Talento Humano, Tecnología, Planeación y Control, Comunicaciones, Investigaciones Económicas y Cumplimiento (Corficolombiana, 2025c).

Figura 1

Esquema Simplificado de la Estructura Organizacional de Corficolombiana



Nota. Elaboración propia, adaptado de Organigrama General, por Corficolombiana, 2025

La Vicepresidencia de Tesorería tiene especial relevancia para esta investigación. Bajo su estructura se ubican la Mesa ALM, la Mesa de Moneda Extranjera, la Mesa de Moneda Legal, la Mesa Institucional y el frente de derivados, lo que evidencia que la organización dispone de capacidades diferenciadas para administrar liquidez, riesgo de tasa, riesgo cambiario, operaciones de mercado monetario y relación con clientes institucionales y corporativos (Corficolombiana, 2025c). Esta estructura hace viable una intervención focalizada en pricing de CDTs, pues las decisiones de captación y fondeo dialogan de manera natural con la lectura de curvas, las condiciones de liquidez y la estrategia comercial de tesorería (Corficolombiana, 2025b).

La existencia de una Vicepresidencia de Gobierno, Riesgo y Cumplimiento también es significativa para la propuesta. En un proyecto de ciencia de datos aplicado a un proceso financiero sensible, la gobernanza del modelo, la calidad del dato y la consistencia de los controles son elementos tan importantes como la capacidad predictiva (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011). Por ello, la estructura organizacional de Corficolombiana no solo soporta la ejecución técnica del proyecto, sino también su eventual adopción bajo criterios de riesgo de modelo, control interno y cumplimiento regulatorio.

Productos y Servicios

El portafolio de Corficolombiana integra negocios de tesorería, inversiones, banca de inversión y servicios financieros complementarios (Corficolombiana, s/f-b). En el ámbito más próximo a esta investigación, la entidad participa en operaciones de contado, mercado monetario, títulos de deuda pública y privada, derivados y captaciones. Su portafolio de productos y servicios de 2025 confirma que Corficolombiana ofrece operaciones de tesorería y mercado de capitales orientadas tanto a gestión propia como a atención de clientes corporativos e institucionales, lo que incluye compraventa de divisas, estructuración de coberturas y operaciones de financiación y fondeo (Corficolombiana, s/f-b).

Para el proyecto, la categoría más relevante es la asociada a captación y manejo del balance. El informe de gestión de 2024 señala que el fondeo mediante CDT mantuvo un peso importante dentro de la estrategia financiera de la entidad y que la tasa efectiva promedio de costo de estos instrumentos siguió siendo una variable crítica para la tesorería (Corficolombiana, 2025b). En consecuencia, el pricing de CDTs debe entenderse no solo como una decisión comercial de tasas, sino como un componente de la administración integral del costo de fondeo, del margen financiero y de la competitividad frente al mercado.

Posicionamiento en el Mercado

Corficolombiana mantiene una posición relevante en segmentos de renta fija, divisas y derivados, lo que fortalece la pertinencia del caso de estudio. El informe CorfiSostenible 2024 reporta que la entidad continuó como uno de los participantes líderes en el mercado local de deuda pública y que, al cierre de 2024, ocupó el puesto 10 dentro del esquema de Creadores de Mercado del Ministerio de Hacienda y Crédito Público (Corficolombiana, 2025b). En el mercado spot peso-dólar, la organización alcanzó una participación de 16,95 %, mientras que en forwards registró una participación de 0,98 %. El portafolio de derivados forward peso-dólar cerró en USD 2.792 millones y el libro de opciones en USD 423 millones, cifras que reflejan una operación de tesorería material y compleja (Corficolombiana, 2025b).

Sin embargo, el posicionamiento de mercado convive con un reto financiero concreto: el costo del fondeo. El mismo informe de gestión indica que la tesorería de Corficolombiana cerró 2024 con pérdidas por COP 56.453 millones, asociadas principalmente al carry negativo derivado del elevado costo del financiamiento, y señala que en diciembre de 2024 la tasa efectiva promedio de costo en CDTs fue de 11,01 % (Corficolombiana, 2025b). Aunque el reporte de resultados del primer trimestre de 2025 mostró una mejora frente al mismo periodo de 2024, el margen neto de tesorería permaneció en

terreno negativo, lo que sugiere que el problema sigue vigente y requiere una respuesta más estructurada (Corficolombiana, 2025d).

En consecuencia, Corficolombiana cuenta con la escala, la complejidad operativa y la relevancia de mercado suficientes para que una intervención en pricing de CDTs tenga sentido estratégico. No se trata de una entidad marginal ni de un proceso accesorio, sino de una corporación financiera con exposición real a tasas, curvas, competencia y costo de fondeo, cuya estructura institucional permite tanto la implementación técnica como la futura gobernanza de una solución analítica de apoyo a la decisión.

Ubicación del proyecto dentro de la operación de Corficolombiana

En términos operativos, el proyecto se inserta en la Vicepresidencia de Tesorería, específicamente en el proceso de definición de tasas y spreads de captación para CDTs, en articulación con las funciones de captación, gestión de fondeo, ALM y mesas de dinero. Su aporte se ubica en la capa de apoyo analítico a la decisión comercial, al integrar señales de mercado, variables macroeconómicas y referencias competitivas para fortalecer la consistencia, trazabilidad y comparabilidad del pricing. En consecuencia, la propuesta no interviene la totalidad de la operación de Corficolombiana, sino un proceso puntual y estratégico relacionado con el costo de fondeo y la administración del balance.

Marco de Referencia

Certificados de Depósito a Término (CDTs): instrumento, modalidades e indexadores

El Certificado de Depósito a Término (CDT) es un instrumento de captación mediante el cual una entidad financiera recibe recursos del público por un plazo definido y se obliga a devolver el capital junto con un rendimiento acordado al vencimiento. El Banco de la República lo describe como una inversión a plazo determinado cuya tasa de captación depende del monto y del plazo pactados (Banco de la República, s/f-a) mientras que la Superintendencia Financiera de Colombia (SFC) subraya que sus rasgos jurídicos esenciales son la irredimibilidad antes del vencimiento, la libre negociación en el mercado secundario y el carácter nominativo del título (SFC, s/f).

Desde el punto de vista del pricing, los CDTs en Colombia se clasifican en función del indexador o referencia sobre la cual se expresa la tasa acordada (Banco de la República, s/f-a). Las principales modalidades son cuatro. La primera es la tasa fija, en la que el rendimiento se expresa como un porcentaje nominal anual independiente de cualquier variable de referencia, lo que la convierte en la modalidad más directamente comparable contra la curva soberana de TES (SFC, 2006). La segunda es la indexación al Indicador Bancario de Referencia (IBR), en la cual la tasa se pacta como un spread sobre el IBR overnight o a distintos plazos, reflejando así las condiciones del mercado monetario de corto plazo; como señala el Banco de la República, el IBR es una tasa de referencia del mercado monetario colombiano que se calcula diariamente con base en las cotizaciones de un panel de entidades financieras (Banco de la República, s/f-b). La tercera modalidad es la indexación al Índice de Precios al Consumidor (IPC), en la cual la tasa pactada corresponde a un spread sobre la inflación anualizada certificada por el DANE, lo que traslada al depositante la protección contra la erosión del poder adquisitivo y le exige a la entidad gestionar el riesgo de base entre inflación realizada y expectativas (DANE, 2025a). La cuarta modalidad es la indexación a la Unidad de Valor Real (UVR), una unidad de cuenta actualizada diariamente por el Banco de la República con base en la inflación del IPC del mes

anterior (Banco de la República, s/f-e) utilizada principalmente en operaciones de mayor plazo o con clientes institucionales especializados.

La existencia de estas cuatro modalidades hace que el pricing de CDTs no sea un problema de tasa única, sino un problema de superficie de tasas: para cada combinación de plazo e indexador existe una referencia de mercado distinta, una curva de comparación diferente y un spread que debe calibrarse a las condiciones del momento. Esta complejidad es precisamente la que justifica la construcción de un motor prescriptivo basado en datos, en lugar de una definición ad hoc de tasas.

En cuanto al impacto macroeconómico sobre el pricing, tres variables son especialmente relevantes para el caso colombiano. La tasa de política monetaria del Banco de la República fija el piso de referencia del costo de fondeo del sistema; sus cambios se transmiten con rezago a las tasas de captación y condicionan el nivel general de la curva (Banco de la República, s/f-d). El IPC define tanto el rendimiento real esperado de los CDTs indexados como el contexto de poder adquisitivo de los depositantes. La Tasa Representativa del Mercado (TRM), que expresa el precio de equilibrio del dólar frente al peso en el mercado spot (Banco de la República, s/f-c), afecta indirectamente el pricing mediante su impacto sobre la inflación importada y los flujos de capitales que determinan las condiciones de liquidez del sistema financiero.

Precia y los datos de mercado en Colombia

Precia S.A. es la sociedad administradora de precios para la valoración de portafolios en Colombia, autorizada por la Superintendencia Financiera. Su función central es calcular y difundir los precios justos de mercado de los instrumentos financieros que negocian los fondos de inversión colectiva, las compañías de seguros, los fondos de pensiones y demás intermediarios del mercado colombiano, de manera que la valoración de portafolios sea homogénea, objetiva y consistente con las condiciones observadas en el mercado secundario (Precia, s/f).

Para efectos de este proyecto, Precia constituye la fuente primaria de datos de mercado, y su arquitectura de archivos estructurados es el punto de partida del pipeline analítico. Los archivos más relevantes son los siguientes. Los archivos de Precia se clasifican en cuatro grupos principales. Los archivos SX contienen información de valoración de instrumentos individuales, incluyendo TES y CDTs, con variables como tasa, TIR, margen, duración, convexidad, calificación, fechas y emisor. Los archivos SC contienen curvas cero cupón por fecha, curva y plazo. Los archivos SB contienen parámetros de curvas, como Beta 0, Beta 1, Beta 2 y Beta 3, asociados a la forma de la estructura temporal. Finalmente, los archivos SwapCC contienen curvas swap, utilizadas como referencia para instrumentos indexados al IBR. Esta clasificación permite transformar archivos planos de mercado en insumos estructurados para el cálculo de spreads, basis y variables explicativas del modelo. La disponibilidad histórica y la estructura estandarizada de estos archivos hacen posible la construcción reproducible de series comparables por plazo e indexador, que es el insumo fundamental para el cálculo del basis de emisión y para el entrenamiento de los modelos de aprendizaje automático.

Curvas de mercado, estructura temporal de tasas y basis de emisión

Estructura temporal de tasas

La estructura temporal de tasas de interés (ETTI) describe la relación entre el rendimiento de instrumentos de renta fija y sus plazos de vencimiento, bajo el supuesto de que el resto de las características —riesgo de crédito, liquidez y moneda— son equivalentes (Andrada Félix et al., 2014). Su representación más utilizada es la curva cero cupón, que expresa tasas spot libres de riesgo para cada horizonte y permite comparar de manera homogénea instrumentos con distintos perfiles de flujos de caja. El BIS documenta que las curvas cero cupón suelen estimarse a partir de instrumentos observables en el mercado secundario mediante métodos de ajuste paramétrico, entre los cuales el modelo de Nelson y Siegel (1987) y su extensión de Svensson (1994) son los más ampliamente empleados por bancos centrales y participantes del mercado alrededor del mundo (BIS, 2005).

El modelo de Nelson-Siegel expresa la tasa forward instantánea como función de tres parámetros que capturan el nivel, la pendiente y la curvatura de la curva, con un parámetro adicional de escala temporal. Su extensión de Svensson incorpora un cuarto término que añade flexibilidad para representar curvas con dos jorobas, característica frecuente en mercados emergentes con alta volatilidad (Alfaro, 2009). En Colombia, Precia utiliza estos parámetros para publicar diariamente las curvas de referencia del sistema, lo que permite a los participantes del mercado construir tasas de comparación para cualquier plazo de manera consistente.

Sin embargo, para un pipeline de datos que debe operar con plazos de emisión no siempre coincidentes con los nodos de la curva publicada, la interpolación es un paso indispensable. En este proyecto se utiliza interpolación PCHIP (Piecewise Cubic Hermite Interpolating Polynomial), una técnica monotónica shape-preserving que garantiza que los valores interpolados no oscilen ni introduzcan artefactos entre nodos conocidos (Le Floc', 2013). Le Floc' (2013) demuestra que la estabilidad de la interpolación en curvas de tasas es un criterio crítico para la valoración, ya que interpolaciones no monotónicas pueden generar tasas forward negativas en segmentos de la curva donde las tasas son crecientes (Le Floc', 2013). La curva TES continua construida con PCHIP en este proyecto sirve como referencia soberana para todos los cálculos de spread y de basis de emisión.

Curva swap y referencia monetaria

La curva swap de tasa de interés expresa el costo de intercambiar flujos de tasa fija por flujos de tasa variable (referenciados al IBR en el caso colombiano) a distintos horizontes. Su importancia para el pricing de CDTs indexados a IBR radica en que permite establecer cuánto costaría reemplazar el fondeo vía CDT por una operación de mercado de capitales de plazo equivalente. Cuando una entidad emite un CDT a IBR + margen, el spread negociado con el cliente puede compararse directamente contra el nivel de la curva swap para ese plazo, determinando así si la captación es más o menos costosa que el fondeo

alternativo en el mercado de derivados. Esta comparación es uno de los principales casos de uso del pipeline descrito en este trabajo.

Basis de emisión como medida relativa de pricing

Uno de los ejes conceptuales del pipeline analítico desarrollado en este trabajo es el basis de emisión, definido como el diferencial entre la condición financiera observada en una emisión de CDT y la referencia de mercado de un instrumento comparable en plazo y estructura de indexación. Bajo este enfoque, el basis representa el componente relativo de la emisión frente al mercado y permite aislar el efecto propio del emisor y de las condiciones específicas de colocación respecto de la curva de referencia vigente.

En el caso de los CDT a tasa fija, este diferencial se mide frente a una referencia de mercado equivalente para ese plazo. En los CDT indexados, la metodología distingue entre el componente base asociado al indexador y el spread de emisión, por lo que la modelación se realiza sobre el spread y no sobre la tasa total observada. Así, tanto para emisiones referidas a IBR como para emisiones referidas a IPC, el basis refleja el exceso o descuento del spread emitido frente a la referencia de mercado correspondiente, haciendo posible una evaluación consistente del pricing relativo de cada colocación.

Un basis positivo indica que la entidad está pagando por encima de lo que el mercado exigiría para un instrumento libre de riesgo del mismo plazo, es decir, está pagando por encima de la referencia de mercado. Un basis nulo o negativo sugiere una captación más eficiente relativa al mercado. Este concepto, análogo al credit spread en el mercado de bonos corporativos (BIS, 2015), convierte el precio absoluto de la captación en una medida relativa y comparable entre emisiones de distinto plazo, indexador y momento de mercado, resolviendo así el problema de heterogeneidad que impide comparar tasas de CDTs directamente sin homologación previa. El basis de emisión es una de las

variables objetivo del proyecto, específicamente en el escenario BASIS_GENERAL. Los demás escenarios utilizan targets diferenciados: TIR para tasa fija y Spread_Indexado para instrumentos IBR e IPC.

Para efectos metodológicos, este trabajo distingue entre el basis de emisión como concepto financiero general y la prima inicial calculada operacionalmente en el pipeline. En términos conceptuales, el basis corresponde a la diferencia entre la tasa o spread observado de una emisión y una referencia de mercado comparable por plazo, indexador y calidad crediticia. Sin embargo, en la primera etapa del pipeline, la variable Prima_Emision_Basis se calcula a partir de la información inicial observada en Precia: para CDTs a tasa fija, como la diferencia entre la TIR observada y la tasa facial; y para instrumentos indexados, como la diferencia entre el margen observado y la tasa o spread registrado. Posteriormente, en la etapa de feature engineering, esta prima inicial se complementa con variables construidas frente a curvas soberanas, curvas AAA, curva swap y referencias competitivas, permitiendo una lectura más completa del posicionamiento relativo de cada emisión.

Gestión de activos y pasivos (ALM), riesgo de tasa de interés y Precio de Transferencia de Fondos

La gestión de activos y pasivos (ALM) es la disciplina que busca optimizar el balance de una entidad financiera controlando simultáneamente el riesgo de tasa de interés, el riesgo de liquidez y el riesgo de base. El Comité de Basilea (BCBS, 2016) define el Riesgo de Tasa de Interés del Libro Bancario (RTILB) como la exposición del balance estructural a movimientos adversos de las tasas, que afectan tanto los ingresos netos por intereses como el valor económico del patrimonio. En Colombia, Asobancaria ha señalado que el RTILB refleja la vulnerabilidad del balance ante cambios en las tasas y que sus principales subtipos son el riesgo de brecha, el riesgo de base entre indexadores y el riesgo de opciones embebidas en productos de captación (Asobancaria, 2020).

Para efectos de este trabajo, el marco ALM y el RTILB son relevantes por dos razones concretas. Primero, porque el pricing de CDTs no es una decisión comercial aislada: cada tasa que se ofrece afecta el costo de fondeo del balance y, por ende, el margen neto de interés de la entidad. Una tasa demasiado alta deteriora el margen; una tasa demasiado baja puede comprometer la competitividad comercial y el volumen captado. Esta tensión es precisamente la que el motor prescriptivo busca apoyar con información estructurada. Segundo, porque la Circular Externa 025 de 2022 de la SFC instruye a los establecimientos de crédito colombianos a implementar metodologías formales de medición y control del RTILB, lo que hace que la trazabilidad y la documentación del proceso de pricing cobren relevancia regulatoria adicional (SFC, 2022b).

El Precio de Transferencia de Fondos (FTP) es la herramienta mediante la cual la tesorería de una entidad financiera establece la tasa interna a la que compra los recursos captados por las mesas de fondeo y los vende a las unidades de colocación (KPMG, 2016). KPMG (2016) describe el FTP como el mecanismo que centraliza los riesgos de tasa, liquidez y divisa en una única unidad de gestión (el Centro de Finanzas, o CFU), que cuenta con los recursos y la especialización para medir y cubrir esos riesgos (KPMG, 2016). La curva FTP se construye sobre una referencia de mercado —típicamente la curva swap o la curva soberana— a la que se adicionan primas por liquidez a plazo, costos de reserva de liquidez contingente y ajustes por franquicia de depósitos (Moody's, 2021). Para el presente proyecto, el FTP es una de las señales más citadas por los informantes clave como criterio de validación de las tasas propuestas, pues permite verificar si una captación a una tasa determinada es financieramente sostenible para el balance de la entidad, no solo comercialmente competitiva.

Respecto a los depósitos sin vencimiento (NMD), que son los depósitos a la vista —cuentas de ahorro y corrientes— sin fecha contractual de vencimiento, cabe mencionarlos brevemente como contexto del marco ALM, dado que conforman otra fuente de fondeo que convive con los CDTs en el balance de la organización. A diferencia de los CDTs, los NMD requieren modelación comportamental

para estimar su vida efectiva y su sensibilidad a las tasas, pues los clientes pueden retirarlos en cualquier momento (van Dyk, 2025). Sin embargo, esta característica los convierte en un problema de modelación distinto al del pricing de CDTs, que tienen plazo definido contractualmente. Por ello, aunque el contexto ALM los menciona como parte del marco de gestión del balance, no constituyen el objeto de esta investigación ni una variable central del pipeline analítico.

Ciencia de datos aplicada al pricing financiero: pipelines, ingeniería de datos y modelado

El pipeline de datos como artefacto analítico

Un pipeline de datos es una secuencia automatizada y reproducible de pasos que transforma datos crudos provenientes de distintas fuentes en información lista para el análisis o la toma de decisiones. En el contexto de la ciencia de datos aplicada al sector financiero, el pipeline no es un componente accesorio del proyecto, sino su columna vertebral: es el mecanismo que garantiza que los resultados analíticos sean reproducibles, auditables y actualizables cuando cambian las condiciones del mercado. Tal como señala la guía CRISP-DM (Chapman et al., 2000), las fases de comprensión y preparación de los datos consumen habitualmente más del 60 % del esfuerzo total de un proyecto de minería de datos, y su calidad determina directamente la confiabilidad del modelo resultante.

En términos técnicos, un pipeline de datos estructurado para análisis financiero comprende al menos cuatro capas funcionales. La primera es la capa de ingesta y persistencia, responsable de adquirir los datos de sus fuentes originales —archivos planos, APIs, bases de datos— y almacenarlos en formatos optimizados para lectura analítica. En este proyecto, los datos de Precia se descargan en archivos estructurados que se procesan mediante parsers específicos y se persisten en formato Apache Parquet. Parquet es un formato columnar de almacenamiento binario ampliamente utilizado en ciencia de datos porque permite lectura selectiva de columnas, compresión eficiente y carga rápida en herramientas de análisis como pandas y Apache Spark (Apache Software Foundation, 2024). La segunda capa es la de

transformación y homologación, donde los datos crudos se limpian, se homologan —en este caso, plazos en distintas unidades se convierten a días calendario, tasas en distintas bases se anualizan, e indexadores distintos se reconcilian— y se integran en estructuras analíticas coherentes. La tercera capa es la de construcción de variables, o *feature engineering*, donde se crean las variables explicativas que alimentarán los modelos. La cuarta capa es la de modelado y despliegue, donde los algoritmos de aprendizaje automático se entrenan, se evalúan y sus salidas se traducen en recomendaciones para el usuario.

Ingeniería de características financieras (Feature Engineering)

La ingeniería de características es el proceso de construir variables explicativas a partir de los datos crudos mediante transformaciones matemáticas, combinaciones de variables o incorporación de contexto de dominio. En modelos de pricing financiero, las características más valiosas no son siempre las variables en bruto, sino las transformaciones que capturan relaciones económicas relevantes: spreads entre curvas, diferencias entre plazos, variaciones porcentuales, promedios móviles y rezagos temporales (Hastie et al., 2009).

Para el pipeline de este proyecto, las características se construyen en tres grupos. El primero son las características de emisión: plazo en días, indexador codificado, entidad emisora y monto de la emisión, que describen las condiciones básicas del instrumento. El segundo son las características de mercado al momento de la emisión: nivel de la curva TES al plazo correspondiente interpolada con PCHIP, nivel de la curva swap IBR para ese plazo, spread entre la curva AAA corporativa y la soberana, nivel de la tasa de política monetaria del Banco de la República, e IPC anualizado del período. El tercero son las señales competitivas: medidas resumidas del comportamiento de captación de entidades comparables, derivadas del histórico de emisiones del sistema disponible en los archivos SX de Precia. La combinación de estas tres familias de características permite que el modelo aprenda no solo el nivel

absoluto de las tasas, sino también cómo se posiciona cada emisión relativa al mercado en un momento dado, que es exactamente la lógica del basis de emisión.

Un control crítico en el feature engineering de modelos de pricing financiero es la prevención del data leakage o fuga de información, que ocurre cuando se incluyen en el conjunto de entrenamiento variables que contienen información del futuro o del target en sí misma (Hastie et al., 2009). En este proyecto, el control de leakage implica excluir variables que son consecuencia de la decisión de pricing —como el volumen captado después de la emisión— y asegurarse de que todas las señales de mercado utilizadas como características estaban efectivamente disponibles en el momento de la decisión de fijación de tasa.

Algoritmos de aprendizaje automático supervisado para regresión

El aprendizaje automático supervisado es la rama de la inteligencia artificial en la que un modelo aprende la relación entre un conjunto de variables de entrada (características) y una variable de salida continua (el target) a partir de ejemplos históricos etiquetados, con el objetivo de predecir el valor de la variable de salida para nuevas observaciones (James et al., 2023). En este proyecto, la variable objetivo no se define como un único target universal. La modelación se estructuró por escenarios de pricing: para los CDTs a tasa fija se modeló la TIR; para los CDTs indexados al IBR se modeló el spread indexado; para los CDTs indexados al IPC se modeló también el spread indexado; y, de manera complementaria, se modeló la prima o basis de emisión en un escenario general. Esta separación permite respetar la lógica financiera de cada modalidad de CDT, dado que una emisión a tasa fija no se comporta igual que una emisión indexada a IBR o IPC.

Se exploraron los siguientes algoritmos en el torneo de modelos del proyecto:

Regresión Ridge y Lasso. Son extensiones del modelo de regresión lineal que incorporan términos de penalización (regularización) para reducir el sobreajuste cuando el número de

características es elevado en relación con las observaciones disponibles. Ridge (regularización L2) encoge los coeficientes hacia cero sin eliminarlos, mientras que Lasso (regularización L1) puede llevar coeficientes a exactamente cero, realizando selección de variables implícita (Tibshirani, 1996). En el contexto financiero, estos modelos son atractivos por su interpretabilidad directa: cada coeficiente expresa el cambio esperado en el basis por cada unidad de cambio en la variable correspondiente.

ElasticNet. Combina las penalizaciones L1 y L2 en un mismo modelo lineal regularizado. Esta combinación permite equilibrar dos objetivos: reducir la magnitud de los coeficientes cuando existen variables correlacionadas y, al mismo tiempo, conservar cierta capacidad de selección de variables. En contextos financieros con curvas, spreads, tasas y señales competitivas altamente relacionadas, ElasticNet resulta útil porque ofrece estabilidad, control de sobreajuste e interpretabilidad superior frente a modelos no lineales más complejos (Zou & Hastie, 2005).

Random Forest. Es un método de ensamble que construye un gran número de árboles de decisión sobre submuestras aleatorias de los datos y del espacio de características, y promedia sus predicciones para reducir la varianza y aumentar la robustez (Breiman, 2001). Los árboles de decisión individuales aprenden umbrales de las variables que dividen el espacio de características en regiones con valores similares del target; el promedio sobre cientos o miles de árboles suaviza la predicción y la hace más estable. Los Random Forests son especialmente útiles cuando existen interacciones no lineales entre las variables explicativas, como la interacción entre el plazo de la emisión y el nivel de la curva de tasas, que puede determinar si una prima de liquidez es adecuada o no.

Gradient Boosting (XGBoost, LightGBM). Son métodos de ensamble secuenciales en los que cada árbol sucesivo se ajusta para corregir los errores residuales del árbol anterior, optimizando directamente una función de pérdida diferenciable (Friedman, 2001). XGBoost y LightGBM (Ke et al., 2017) son implementaciones altamente eficientes de gradient boosting que incorporan regularización,

manejo de valores faltantes y paralelización, y que han demostrado un desempeño superior en una amplia variedad de tareas de regresión con datos tabulares (Chen & Guestrin, 2016). En el contexto del pricing de CDTs, el gradient boosting puede capturar relaciones complejas entre el ciclo de política monetaria, el plazo y el nivel de la competencia que los modelos lineales no representan fácilmente.

Adicionalmente, se evaluó un esquema de Stacking, técnica de ensamble que combina las predicciones de varios modelos base mediante un meta-modelo final (Wolpert, 1992). En este proyecto, el Stacking se utilizó como alternativa comparativa dentro del torneo de modelos, combinando XGBoost, LightGBM y Random Forest como modelos base, con Ridge como estimador final. Su inclusión permitió evaluar si una arquitectura de ensamble aportaba mayor precisión que los modelos individuales; sin embargo, los resultados mostraron que no fue el modelo ganador en ninguno de los escenarios.

Evaluación y selección de modelos

La evaluación de modelos de regresión se realiza mediante métricas de error que cuantifican la distancia entre los valores predichos y los valores reales del target en un conjunto de datos no visto durante el entrenamiento. Las métricas utilizadas en este proyecto son:

RMSE (Root Mean Squared Error): raíz cuadrada del error cuadrático medio, que penaliza los errores grandes con mayor intensidad y tiene las mismas unidades que el target.

MAE (Mean Absolute Error): error absoluto medio, interpretable directamente como la desviación promedio entre la predicción y el valor real.

R^2 (coeficiente de determinación): proporción de la varianza del target explicada por el modelo, donde 1 indica predicción perfecta y 0 indica que el modelo no mejora sobre el promedio del target.

La selección del modelo definitivo (champion model) no descansa únicamente en el menor error, sino también en la plausibilidad económica de sus predicciones, su estabilidad ante distintos

subperiodos de evaluación y su capacidad de explicación gerencial. Este criterio dual —técnico y económico— es consistente con las recomendaciones de la guía SR 11-7 sobre validación de modelos en entidades financieras (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011).

Para garantizar la validez de la evaluación, se utiliza una separación temporal entre el conjunto de entrenamiento y el conjunto de prueba (train-test split cronológico), de modo que el modelo no pueda ver datos futuros durante su entrenamiento, replicando las condiciones reales de uso en producción.

Interpretabilidad de modelos de aprendizaje automático

Un modelo de machine learning solo es útil en un contexto financiero regulado si sus decisiones pueden explicarse a usuarios de negocio, equipos de riesgo y auditores. La interpretabilidad se refiere a la capacidad de un modelo para exponer de manera comprensible las razones que conducen a una predicción particular, ya sea a nivel global (qué variables importan en promedio) o a nivel local (qué factores explican una predicción específica) (Molnar, 2019).

El enfoque más utilizado para la interpretabilidad global de modelos de ensamble es la importancia de características (feature importance), que mide la contribución promedio de cada variable a la reducción del error durante el entrenamiento. En Random Forest y Gradient Boosting, la importancia se calcula como la reducción total de impureza (o pérdida) atribuible a cada característica sumada sobre todos los árboles y nodos del ensamble (Breiman, 2001).

Para la explicabilidad local —es decir, para explicar por qué el modelo recomienda una tasa específica para una emisión concreta— el enfoque más sólido desde el punto de vista teórico son los valores SHAP (SHapley Additive exPlanations), basados en la teoría de valor de Shapley de la teoría de juegos cooperativos. Los valores SHAP asignan a cada característica una contribución marginal a la

predicción del modelo para una observación dada, de manera que la suma de todas las contribuciones más el valor base del modelo reproduce exactamente la predicción (Lundberg et al., 2017). En el contexto de este proyecto, los valores SHAP permiten responder preguntas del tipo: "¿cuánto de la tasa recomendada de este CDT se debe al nivel actual de la curva TES, y cuánto se debe al spread de la competencia?". Esta capacidad de descomposición es exactamente lo que los informantes clave del diagnóstico identificaron como necesidad crítica para confiar en la herramienta.

Aunque SHAP constituye una referencia robusta para explicabilidad local, la versión implementada del prototipo se concentró en interpretabilidad global y diagnóstica mediante importancia por permutación, coeficientes agregados, importancia nativa, análisis de residuales, matrices de correlación y matrices de deciles. La incorporación de SHAP se plantea como una extensión futura para explicar recomendaciones individuales.

La importancia de la interpretabilidad en modelos financieros ha sido reconocida por los reguladores. La guía SR 11-7 (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011) establece que los modelos utilizados en decisiones financieras relevantes deben ser suficientemente comprensibles para quienes los utilizan y para quienes los validan. La Financial Stability Board (FSB, 2024) y el NIST AI RMF (Tabassi, 2023) van más allá y señalan que la explicabilidad es un componente del uso responsable de la inteligencia artificial, no solo un requisito de comunicación.

Gobierno de datos en el sector financiero

Definición y principios

El gobierno de datos es el conjunto de políticas, procesos, estándares y roles que aseguran que los datos de una organización sean precisos, consistentes, trazables, seguros y utilizados de manera apropiada para la toma de decisiones (DAMA International, 2017). En el sector financiero, el gobierno de

datos no es una práctica optativa sino un requisito regulatorio, dado que las decisiones que dependen de datos —desde la valoración de portafolios hasta la fijación de tasas de captación— deben ser auditables y reproducibles.

El estándar internacional de referencia para el gobierno de datos en entidades financieras es el BCBS 239, publicado por el Comité de Basilea bajo el nombre de Principios para una efectiva agregación de datos de riesgo y reporte de riesgos (BCBS, 2013a). Este estándar establece catorce principios organizados en cuatro dominios: gobernanza y arquitectura de datos, capacidades de agregación de datos de riesgo, prácticas de reporte de riesgos, y revisión supervisora. Los principios más directamente relevantes para este proyecto son los relacionados con la precisión e integridad de los datos (Principio 3), la completitud (Principio 4), la oportunidad (Principio 5) y la adaptabilidad de los reportes (Principio 8). Aunque BCBS 239 fue concebido para la gestión del riesgo sistémico, su lógica se aplica con igual pertinencia a cualquier pipeline analítico que alimenta decisiones financieras: si los datos de entrada son inconsistentes o no trazables, las recomendaciones del motor prescriptivo serán igualmente cuestionables (BCBS, 2013c).

Linaje de datos y trazabilidad

El linaje de datos (data lineage) es la documentación del recorrido que sigue un dato desde su fuente original hasta su uso final, incluyendo todas las transformaciones intermedias. En un pipeline de datos financieros, el linaje responde preguntas fundamentales para la auditoría: ¿de qué archivo proviene esta tasa?, ¿en qué fecha fue descargada?, ¿qué regla de transformación se aplicó para pasar de la tasa nominal publicada al spread calculado?, ¿qué versión del script ejecutó esa transformación? (DAMA International, 2017).

En el contexto de este proyecto, el linaje de datos se materializa en tres artefactos concretos. Las bitácoras de ejecución registran la fecha, el archivo de origen, el número de registros procesados y

cualquier excepción durante la ingesta. Los archivos Parquet versionados preservan estados intermedios del pipeline que pueden consultarse retrospectivamente. El diccionario de datos documenta para cada variable su definición, su unidad, su rango esperado, su fuente y las reglas de transformación aplicadas. Esta infraestructura documental es la que permite reconstruir una recomendación de pricing de meses anteriores, respondiendo exactamente a la necesidad identificada por los informantes clave en el diagnóstico del proceso actual.

Calidad del dato

La calidad del dato es la condición que hace que un dato sea apto para el uso al que se destina, y se evalúa típicamente en dimensiones como precisión (¿el valor es correcto?), completitud (¿no faltan registros?), consistencia (¿los mismos valores se representan de la misma manera en distintas fuentes?), oportunidad (¿el dato está disponible cuando se necesita?) y unicidad (¿no hay duplicados?) (DAMA International, 2017). En pipelines de datos de mercado como el de este proyecto, los problemas de calidad más frecuentes son los valores faltantes en días de baja negociación, los cambios de convención en la expresión de plazos entre versiones de los archivos de Precia, y las discontinuidades en la serie cuando cambian los parámetros de valoración de curvas entre períodos.

La detección y corrección de problemas de calidad debe realizarse de manera sistemática y documentada, no ad hoc. Las reglas de calidad deben estar explícitas en el código del pipeline y producir alertas cuando un dato no cumple el criterio esperado, de modo que la decisión de imputar, excluir o escalar el problema quede registrada y sea reproducible (BCBS, 2013a).

Riesgo de modelo y gobierno del modelo en entidades financieras

El riesgo de modelo es la posibilidad de que las decisiones tomadas con base en un modelo resulten incorrectas debido a errores en su conceptualización, sus datos de entrada, su calibración o su uso (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency,

2011). La guía SR 11-7 de la Reserva Federal y la OCC establece que todas las entidades financieras que dependen de modelos cuantitativos para la toma de decisiones deben contar con un ciclo completo de gestión del riesgo de modelo que incluya: (i) desarrollo y documentación del modelo, (ii) validación independiente, (iii) gobierno y aprobación, y (iv) seguimiento continuo del desempeño.

La validación del modelo implica verificar que sus supuestos son razonables, que sus datos de entrenamiento son representativos, que su desempeño en datos fuera de muestra es suficiente y que sus salidas son coherentes con el criterio experto del negocio (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011). En el contexto de este proyecto, la validación incluye la separación cronológica entre entrenamiento y prueba, la evaluación en distintos subperiodos de mercado y la revisión de los drivers del modelo con los usuarios de tesorería para verificar la plausibilidad económica de las variables que el algoritmo considera más importantes.

El marco NIST AI RMF 1.0 (Tabassi, 2023) amplía la perspectiva del gobierno del modelo hacia el ciclo de vida completo del sistema de inteligencia artificial, e introduce cuatro funciones de gestión de riesgo: Gobernanza (GOVERN), Mapeo de riesgos (MAP), Medición (MEASURE) y Gestión (MANAGE). Para un motor prescriptivo de pricing, estas funciones se traducen en: definir roles y responsabilidades sobre la herramienta, identificar los escenarios en los que el modelo podría fallar, medir el desempeño de forma continua y ajustar o recalibrar cuando el desempeño se deteriore por cambios en las condiciones de mercado.

La Financial Stability Board (FSB, 2024) ha advertido específicamente que el uso de inteligencia artificial en el sector financiero introduce riesgos de concentración cuando muchas entidades usan los mismos modelos, riesgos de opacidad cuando los modelos no son explicables y riesgos de amplificación cuando los modelos reaccionan de manera procíclica ante cambios de mercado. Por ello, la FSB recomienda que las entidades que adopten IA en funciones financieras cuenten con mecanismos de

supervisión humana, límites claros al uso autónomo de los modelos y procesos de revisión periódica. Estos principios son plenamente aplicables al motor prescriptivo propuesto en este trabajo y se reflejan explícitamente en el plan de intervención formulado.

Marco normativo colombiano relevante para el pricing de CDTs

El proceso de pricing de CDTs en Colombia opera dentro de un marco normativo que establece obligaciones tanto para la captación de recursos como para la gestión del riesgo de mercado y la calidad de la información. Las regulaciones más directamente relevantes para este proyecto son las siguientes.

La Circular Básica Contable y Financiera (CBCF) de la Superintendencia Financiera establece los criterios de valoración a precios de mercado de los portafolios de las entidades vigiladas y define el papel de los proveedores de precios como Precia en ese proceso (SFC, 2022a). La Circular Externa 025 de 2022 de la SFC exige que los establecimientos de crédito adopten metodologías formales de medición del Riesgo de Tasa de Interés del Libro Bancario (RTILB), establezcan límites de apetito de riesgo y reporten periódicamente las sensibilidades de su NII y EVE ante escenarios de choque de tasas (SFC, 2022a). Esta exigencia refuerza la necesidad de contar con herramientas analíticas trazables para la definición de tasas de captación, dado que el pricing de CDTs es una decisión que afecta directamente las métricas del RTILB.

La Circular Externa 004 de 2024 de la SFC sobre finanzas abiertas establece un marco de interoperabilidad de datos financieros que, a mediano plazo, aumentará la disponibilidad de señales competitivas y de mercado accesibles para los participantes del sistema (Holland & Knight, 2024). En la perspectiva del banco central, el Reporte de Estabilidad Financiera del Banco de la República (2024) advierte que la mayor exposición a títulos de deuda pública incrementa la vulnerabilidad de las entidades ante movimientos adversos en la prima de riesgo soberano, lo que subraya la importancia de

integrar la lectura de la curva soberana en el proceso de pricing de captaciones (Banco de la República, 2024b).

Diseño Metodológico

Este capítulo responde de manera explícita a los requerimientos de la plantilla institucional: enfoque del estudio, fases para alcanzar los objetivos, fuentes y unidad de análisis, instrumentos y validación, procedimientos para el diagnóstico interno y externo, metodología para formular el plan de mejora y ficha técnica (Hernández-Sampieri & Mendoza Torres, 2018). El diseño metodológico no se limita a describir el pipeline analítico, sino que articula el diagnóstico empresarial con la solución prescriptiva propuesta para el pricing de CDTs en Corficolombiana.

El trabajo adopta la lógica propia del trabajo dirigido: una intervención empresarial aplicada, con predominio cuantitativo y soporte documental y cualitativo focalizado, orientada a resolver un problema concreto de negocio (Bernal Torres, 2016). Esta formulación es más consistente con la guía de la universidad, que pide describir la metodología de diagnóstico y de intervención, que encasillar el estudio en un diseño clásico de investigación experimental o no experimental.

Enfoque del estudio y caracterización metodológica

La intervención se orienta a mejorar la decisión de pricing de CDTs en Corficolombiana mediante un motor prescriptivo que integra curvas de mercado, señales macroeconómicas y variables competitivas. Por ello, el predominio metodológico es cuantitativo; sin embargo, el trabajo también incorpora revisión documental, análisis estratégico y juicio experto en la definición de reglas, criterios de gobierno y ruta de implementación (Hernández-Sampieri & Mendoza Torres, 2018). Esta combinación es consistente con la investigación aplicada y con los trabajos dirigidos de carácter empresarial, en los que la utilidad práctica del resultado es tan importante como su rigor analítico (Bernal Torres, 2016)

La unidad de análisis no es una persona ni un área aislada, sino la decisión organizacional de pricing de CDTs: qué tasa o spread ofrecer para un plazo e indexador determinados, bajo condiciones observadas de mercado y de competencia. Desde esta perspectiva, el estudio trabaja sobre registros

históricos y artefactos analíticos del proceso, no sobre percepciones subjetivas como fuente principal. La encuesta no sustituye la capa cuantitativa; la complementa y la triangula, especialmente para el diagnóstico interno, la comprensión de reglas tácitas y la formulación de recomendaciones de adopción (Hernández-Sampieri & Mendoza Torres, 2018).

A continuación, en la siguiente ficha de caracterización metodológica encontramos un resumen de cada dimensión:

Tabla 1

Ficha de Caracterización

Dimensión	Enfoque seleccionado	Justificación
Naturaleza del trabajo	Intervención empresarial aplicada	Se orienta a resolver un problema concreto de pricing y a entregar un artefacto usable para tesorería.
Predominio metodológico	Cuantitativo con apoyo cualitativo focalizado	La evidencia central proviene de series históricas, curvas, variables de mercado y salidas del pipeline; el instrumento a expertos fortalece el diagnóstico interno y la adopción.
Unidad de análisis	Decisión de pricing de CDTs	El foco está en la tasa o spread sugerido y en los factores que explican esa recomendación.
Lógica de validación	Triangulación documental, analítica y experta	Se contrastan fuentes, proceso real, resultados del pipeline y juicios de informantes clave.
Resultado esperado	Motor prescriptivo con trazabilidad y lineamientos de gobierno	La propuesta busca mejorar costo de fondeo, consistencia decisional y auditabilidad.

Nota. Elaboración propia con base en Bernal Torres (2016) y Hernández-Sampieri & Mendoza Torres (2018)

Fases para alcanzar los objetivos del estudio

Las actividades del estudio se organizaron en cinco fases encadenadas. Las dos primeras corresponden a la lógica académica del trabajo dirigido, correspondiente a revisión teórica y diagnóstico

empresarial; las siguientes traducen ese diagnóstico al diseño analítico del pipeline, al modelado/evaluación y a la formulación del plan de intervención. Esta secuencia permite que los objetivos no se queden en actividades sueltas, sino que desemboquen en productos verificables.

Tabla 2

Fases del Diseño Metodológico y Relación Con los Objetivos

Fase	Objetivos a los que aporta	Propósito	Producto principal
1. Revisión de literatura y contexto	OE1	Identificar variables, conceptos y marcos sectoriales, regulatorios y analíticos relevantes para el pricing de captaciones.	Marco teórico y criterios del estudio.
2. Diagnóstico empresarial	OE2	Caracterizar el entorno externo y el proceso actual de pricing mediante análisis estratégico, levantamiento AS-IS y encuesta a informantes clave.	Diagnóstico externo e interno.
3. Diseño del pipeline de datos	OE3	Estructurar la adquisición, persistencia, transformación e integración reproducible de los insumos.	Arquitectura de datos y dataset maestro.
4. Modelado y validación	OE4	Entrenar, comparar e interpretar modelos predictivos y prescriptivos para pricing de CDTs.	Modelos, métricas e interpretabilidad.
5. Formulación del plan de mejora	Transversal	Traducir hallazgos del diagnóstico y del modelado en una propuesta implementable con criterios de adopción y gobierno.	Plan de intervención empresarial.

Nota. Elaboración propia a partir de los objetivos del proyecto.

Metodología de diagnóstico empresarial seleccionada

Como metodología principal para el diagnóstico empresarial se adopta un DOFA estratégico, alimentado por dos herramientas complementarias de diagnóstico externo: las Cinco Fuerzas de Porter y el análisis PESTEL. La elección de este esquema responde a dos razones. Primero, porque la decisión de pricing de CDTs está condicionada simultáneamente por presiones competitivas del sector y por factores

macroeconómicos, tecnológicos y regulatorios (Porter, 2008). Segundo, porque el DOFA permite integrar hallazgos externos e internos en una síntesis orientada a la formulación de estrategias y planes de mejora, no solo a la descripción del entorno (Uribe Macías, 2021).

En términos operativos, las Cinco Fuerzas se utilizan para evaluar rivalidad competitiva, amenaza de sustitutos, poder de clientes, poder de proveedores de liquidez e insumos de mercado, y barreras de entrada relevantes para la captación a plazo (Porter, 2008). El análisis PESTEL se emplea para mapear variables políticas, económicas, sociales, tecnológicas, ambientales y legales que condicionan la formación de tasas y spreads (Uribe Macías, 2021). Finalmente, el DOFA estratégico consolida oportunidades y amenazas del entorno con fortalezas y debilidades internas del proceso de pricing, de modo que el diagnóstico desemboque en lineamientos accionables para la intervención (Uribe Macías, 2021).

El diagnóstico interno, por su parte, se desarrolla a través de cuatro instrumentos complementarios: levantamiento AS-IS del proceso actual, matriz de trazabilidad del pipeline, inventario de fuentes y reglas de negocio, e instrumento de levantamiento a informantes clave. Esta última pieza es importante porque el proceso de pricing no se agota en los datos disponibles: también involucra criterios operativos, secuencias de aprobación, usos reales de la información y fricciones cotidianas que no siempre quedan documentadas. Por ello, el trabajo incorpora una encuesta semiestructurada basada en hechos recientes, siguiendo el principio de preguntar por lo que efectivamente ocurrió y no por lo que las personas creen que debería ocurrir (Fitzpatrick, 2013).

Tabla 3

Herramientas Metodológicas Para el Diagnóstico Externo e Interno

Herramienta	Ámbito	Propósito en el estudio	Salida esperada
-------------	--------	-------------------------	-----------------

Cinco Fuerzas de Porter	Externo	Evaluar intensidad competitiva y presiones de mercado que afectan el pricing de captaciones.	Mapa de rivalidad competitiva e implicaciones para tasas y spreads.
PESTEL	Externo	Caracterizar factores macroeconómicos, tecnológicos y regulatorios relevantes para el negocio de tesorería y fondeo.	Factores del entorno que condicionan la intervención.
DOFA estratégico	Integrador	Sintetizar hallazgos internos y externos para priorizar problemas y orientar la estrategia de mejora.	Matriz de brechas y ejes de intervención.
Levantamiento AS-IS	Interno	Describir el flujo actual de pricing, actores, insumos, aprobaciones y puntos críticos.	Diagnóstico del proceso de pricing y requerimientos funcionales.
Matriz de trazabilidad	Interno	Relacionar fuente, transformación, salida y uso de cada dato del pipeline.	Inventario trazable de insumos
Encuesta a informantes clave	Interno	Recoger evidencia operativa reciente, fricciones del proceso y criterios de adopción	Hallazgos cualitativos y frecuencias descriptivas
Inventario de variables y reglas de negocio	Interno	Verificar coherencia técnica entre objetivos, datos, transformaciones, modelos y salidas.	Base para diseño del pipeline y control de leakage.

Nota. Elaboración propia con base en Porter (2008) y Fitzpatrick (2013)

Población, muestra, instrumentos y validación

La población del componente analítico está conformada por el universo de registros históricos integrados por el pipeline: archivos de mercado Precia, curvas, series del Banco de la República, salidas intermedias y dataset final. Sin embargo, para el componente de diagnóstico interno se define una población distinta: los actores que participan o influyen de manera directa en la decisión de pricing de CDTs dentro de Corficolombiana. Por la naturaleza especializada del proceso y el tamaño acotado del grupo conocedor, se adopta un muestreo intencional o por juicio experto, adecuado cuando se busca profundidad y pertinencia más que representatividad estadística (Campbell et al., 2020).

La muestra prevista para la encuesta–entrevista semiestructurada es de tres a cinco informantes clave. Se priorizan perfiles como trader o responsable de mesa vinculada a fondeo/moneda legal, líder o analista ALM, analista o líder de riesgo con participación en tasas y liquidez, y actor comercial o institucional que participe en la cotización o validación de tasas de captación. El tamaño reducido no constituye una debilidad metodológica, porque el propósito del instrumento no es inferir resultados a una población amplia, sino triangular el diagnóstico documental con evidencia operativa de alto valor contextual.

Tabla 4

Población y Muestra

Elemento	Definición adoptada
Población cualitativa objetivo	Actores directamente vinculados al pricing de CDTs, fondeo, ALM, tesorería y/o riesgo.
Número de personas que intervienen en el proceso de pricing de CDTs	Mesa de moneda legal (2 personas), ALM (2 personas), Riesgo de Liquidez y Mercado (1 persona), y el frente comercial institucional (2 personas)
Muestra prevista	3 a 5 informantes clave.
Técnica de muestreo	Muestreo intencional o por juicio experto.
Criterios de inclusión	Participar en la definición, validación, monitoreo o uso de información relevante para el pricing de CDTs.
Tipo de instrumento	Encuesta–entrevista semiestructurada, con preguntas abiertas, de frecuencia y de priorización.
Propósito del instrumento	Triangular hallazgos del diagnóstico interno y recoger evidencia de hechos recientes del proceso actual.

Nota. Elaboración propia.

Proceso de validación del instrumento

La validación del instrumento se desarrolló en dos etapas, reflejando el proceso iterativo que caracteriza la investigación aplicada. En la primera etapa se elaboró una versión preliminar del

instrumento enfocada en la medición general de las prácticas de pricing y las percepciones sobre gestión de tasas. Esta versión fue sometida a validación de contenido mediante el coeficiente V de Aiken, que cuantifica el nivel de acuerdo entre jueces evaluadores respecto a la pertinencia, claridad y coherencia de cada ítem (Penfield & Giacobbi, 2004).

Los evaluadores, expertos en pricing de CDTs, valoraron los ítems en una escala de 1 a 4. Los resultados mostraron que los ítems con mejor V de Aiken fueron los orientados a prácticas concretas de consulta de fuentes de información y a la documentación de decisiones, mientras que los ítems más generales sobre percepción de eficiencia del proceso obtuvieron valores más bajos por su ambigüedad (Penfield & Giacobbi, 2004). La retroalimentación cualitativa de los jueces señaló además que el instrumento inicial se centraba demasiado en opiniones generales y poco en hechos específicos y recientes, lo que limitaba su utilidad diagnóstica.

En respuesta a estos hallazgos, y en coherencia con el giro metodológico del trabajo hacia una intervención más puntual y prescriptiva, el instrumento fue reformulado en una segunda etapa bajo los principios de The Mom Test (Fitzpatrick, 2013): preguntar sobre lo que efectivamente ocurrió, solicitar ejemplos de la última vez que sucedió algo, evitar hipótesis complacientes y separar opiniones de evidencia conductual. Esta reformulación produjo un instrumento con preguntas más específicas — orientadas a episodios concretos de pricing, fuentes efectivamente consultadas y casos de escalamiento o retrabajo— que resultó más útil para el diagnóstico del proceso real que el instrumento inicial.

La decisión de incluir el V de Aiken en este capítulo responde al valor metodológico que tiene documentar el proceso de validación y ajuste del instrumento, no solo el instrumento final. En investigación aplicada, el ciclo de diseño-validación-mejora del instrumento es evidencia de rigor metodológico y de sensibilidad al contexto organizacional del estudio (Campbell et al., 2020).

Tabla 5*Instrumentos del Estudio y Criterio de Validación*

Instrumento	Uso principal	Criterio de validación	Aplica a
Matriz de inventario de fuentes	Documentar origen, frecuencia, contenido y destino de cada insumo.	Compleitud y trazabilidad fuente-salida.	Diagnóstico y pipeline
Formato Porter–PESTEL–DOFA	Estructurar el diagnóstico externo e integrarlo con hallazgos internos.	Coherencia analítica y pertinencia estratégica.	Diagnóstico organizacional
Matriz AS-IS de proceso	Levantar el proceso actual de pricing y sus puntos críticos.	Correspondencia con el flujo real de trabajo.	Diagnóstico interno
Encuesta–entrevista a informantes clave	Recoger hechos recientes, fricciones y criterios de adopción.	Claridad, pertinencia, coherencia y pilotaje.	Diagnóstico interno
Matriz de variables	Definir objetivos, variables explicativas, targets y reglas de exclusión.	Alineación con objetivos y control de leakage.	Modelado
Ficha técnica de intervención	Sintetizar alcance, unidad de análisis, fuentes, temporalidad y entregables.	Consistencia metodológica del capítulo.	Documento final

Nota. Elaboración propia.

Fuentes de información, procedimientos y técnicas para recoger y analizar la información

Las fuentes utilizadas son predominantemente secundarias y estructuradas: archivos de mercado descargados desde Precia, series del Banco de la República, documentos institucionales de Corficolombiana y scripts desarrollados durante el proyecto. A estas se suma una fuente primaria puntual: las respuestas de informantes clave sobre el proceso actual de pricing. La recolección del componente cuantitativo se realiza mediante descarga, lectura de archivos planos, parseo y consolidación en formato analítico; la recolección del componente cualitativo se realiza mediante aplicación del instrumento en entrevista asistida o formulario acompañado, según disponibilidad de los participantes (Chapman et al., 2000).

Las técnicas de análisis se diferencian por fase. Para el diagnóstico externo se emplean análisis estratégico y lectura estructurada del entorno; para el diagnóstico interno, análisis documental del proceso, matriz AS-IS, revisión de trazabilidad y análisis cualitativo de respuestas abiertas; para la capa analítica, ingeniería de datos, construcción de curvas, feature engineering, modelado supervisado, evaluación comparativa de algoritmos e interpretabilidad (IBM, s/f). Las respuestas abiertas del instrumento se consolidan en una matriz temática, mientras que las preguntas cerradas se resumen con frecuencias y patrones descriptivos, sin recurrir a inferencia estadística debido al tamaño muestral reducido.

Tabla 6

Fuentes de Información y Trazabilidad del Pipeline

Fuente	Contenido principal	Uso en el estudio	Salida asociada
Precia – archivos SX	Valoraciones de TES y CDTs, rendimientos, precios, márgenes y características de emisiones.	Maestro de TES, staging histórico de CDTs y cálculo del panel de basis.	Maestro TES, staging CDT, panel basis
Precia – archivos SC, SB y SwapCC	Curvas cero cupón, betas/parametrizaciones y curvas swap.	Construcción de referencias soberanas, AAA y monetarias para el feature engineering.	Curvas analíticas y spreads
Banco de la República – SDMX	Series macroeconómicas y monetarias relevantes.	Contexto macro y variables explicativas del dataset maestro.	Series macro integradas
Documentos institucionales y sectoriales	Información del proceso actual, entorno competitivo y gobierno aplicable.	Diagnóstico organizacional y lineamientos del plan de mejora.	Matrices diagnósticas y hoja de ruta
Scripts y artefactos del proyecto	Procesos de ingesta, transformación, modelado, interpretabilidad y despliegue.	Trazabilidad metodológica y reproducibilidad técnica.	Parquet, modelos, app, reportes
Encuesta—a expertos	Hechos recientes, usos reales de información y criterios de adopción.	Triangulación del diagnóstico interno.	Hallazgos cualitativos y frecuencias descriptivas

Nota. Elaboración propia a partir de la arquitectura desarrollada y del instrumento propuesto.

Ficha técnica de la intervención

La ficha técnica sintetiza el alcance operativo del estudio y reconoce la coexistencia de dos planos: un universo analítico de registros históricos y un componente cualitativo acotado de informantes clave. Esta combinación es coherente con una intervención empresarial donde la calidad metodológica depende de la trazabilidad, la utilidad práctica y la pertinencia del diagnóstico (Bernal Torres, 2016).

Tabla 7

Ficha Técnica de la Intervención

Característica	Descripción	Fuente o soporte
Tipo de estudio	Intervención empresarial aplicada con predominio cuantitativo y soporte documental.	Diseño metodológico
Objeto de intervención	Proceso de pricing de CDTs y su soporte analítico en Corficolombiana.	Objetivos y diagnóstico
Unidad de análisis	Decisión de pricing expresada como tasa o spread sugerido por plazo e indexador.	Definición del problema
Universo analítico	Registros históricos de mercado, curvas, series macro y salidas del pipeline.	Fuentes integradas
Componente cualitativo	3 a 5 informantes clave vinculados al proceso de pricing, tesorería, ALM, riesgo y/o comercialización institucional.	Instrumento
Ciudad / contexto	Bogotá, Colombia; aplicación en entorno corporativo controlado.	Instrumento
Medio de aplicación del instrumento	Formulario con duración estimada de 20 a 30 minutos.	Instrumento
Técnicas de análisis	Porter, PESTEL, DOFA, ingeniería de datos, modelado supervisado e interpretabilidad.	Capítulo metodológico

Entregables	Diagnóstico, dataset maestro, dataset limpio, modelos, motor prescriptivo, app e intervención propuesta.	Arquitectura del proyecto
Fecha de corte analítico	Ejecución del pipeline con información disponible al 03 de mayo de 2026.	Artefactos técnicos del pipeline y resultados del modelado.

Nota. Elaboración propia.

Variables de estudio, evaluación e interpretabilidad

Las variables objetivo del estudio se definieron por escenario de modelación (IBM, s/f). Para los instrumentos de tasa fija, el target corresponde a la TIR observada. Para los instrumentos indexados al IBR, el target corresponde al Spread_Indexado. Para los instrumentos indexados al IPC, el target también corresponde al Spread_Indexado. Finalmente, para el escenario general, el target corresponde a Prima_Emision_Basis. Esta definición diferenciada evita mezclar instrumentos con lógicas financieras distintas y permite que el motor genere recomendaciones de tasa o spread según el tipo de CDT.

Tabla 8

Variables de Estudio

Tipo	Variable / target	Escenario	Descripción	Uso
Objetivo	TIR	TASA_FIJA	Rendimiento observado de CDTs a tasa fija.	Target para recomendación de tasa fija.
Objetivo	Spread_Indexado	SPREAD_IBR	Rendimiento observado de CDTs indexados al IBR.	Target para recomendación de spread IBR.
Objetivo	Spread_Indexado	SPREAD_IPC	Rendimiento observado de CDTs indexados al IPC.	Target para recomendación de spread IPC.
Objetivo	Prima_Emision_Basis	BASIS_GENERAL	Prima relativa de emisión calculada a partir de información de Precia y enriquecida con referencias de mercado.	Target general de posicionamiento relativo.
Explicativa	Plazo, indexador, emisor y calificación	Todos	Características propias de la emisión.	Segmentación y predicción.

Explicativa	Curvas TES, AAA, swap y PCHIP	Todos	Referencias de mercado por fecha, plazo e indexador.	Construcción de spreads y señales de mercado.
Explicativa	Señales competitivas	Todos	Variables del grupo de bancos comparables.	Medición de presión competitiva.
Control	Variables excluidas por leakage	Todos	Campos que contienen información posterior, identificadores o variables demasiado cercanas al target.	Validez del modelado.

Nota. Elaboración propia.

La evaluación de modelos utiliza RMSE, MAE y R^2 , complementados con criterios de plausibilidad económica y estabilidad. La interpretabilidad implementada en esta versión se aborda mediante importancia por permutación, coeficientes agregados para modelos lineales, importancia nativa para modelos basados en árboles, matrices de correlación, análisis de residuales y matrices de deciles. Estas técnicas permiten identificar drivers globales, evaluar la estabilidad del ajuste y analizar la capacidad de ordenamiento del modelo. La explicabilidad local mediante valores SHAP no se implementó en esta versión y se plantea como una extensión futura para fortalecer la justificación individual de cada recomendación. (Molnar, 2019).

Resultados del instrumento aplicado a informantes clave

El instrumento fue aplicado a cinco informantes clave durante el primer trimestre de 2026, todos vinculados directamente al proceso de pricing de CDTs en Corficolombiana. La totalidad de los participantes otorgó su consentimiento informado y autorizó el uso académico y agregado de sus respuestas. Todos participan en el proceso de pricing con una frecuencia diaria y manejan plazos entre 90 y 1.800 días en modalidades de tasa fija e indexada. La experiencia promedio del grupo es de 22 años en el sector financiero.

Hallazgos cualitativos sobre el proceso actual

Las narrativas de los informantes coinciden en cuatro patrones. Primero, la dependencia del criterio experto individual es alta y reconocida. Todos los participantes describen que la asignación de primas de liquidez para plazos largos, la evaluación de si una señal de competencia es ejecutable o solo indicativa, y la decisión de cuántos puntos básicos mover ante una necesidad puntual de liquidez dependen de la experiencia acumulada del trader o del analista y no de una regla documentada. Como lo expresó uno de los traders: "no existe una regla que me diga: 'si faltan X millones, suba Y puntos básicos'. Depende totalmente de mi experiencia negociando en el día a día".

Segundo, la trazabilidad de la decisión es limitada. Los informantes coinciden en que los sistemas transaccionales registran la tasa final aprobada, pero no las condiciones de mercado que existían en el momento de la decisión. Cuando se han enfrentado a auditorías internas que cuestionan tasas históricas, no siempre fue posible reconstruir con precisión el nivel exacto de las curvas, las cotizaciones competitivas o la situación de liquidez que existía en el momento de la emisión. Uno de los participantes señaló: "en la mesa tomamos decenas de decisiones al día. Mis únicos soportes serían el sistema transaccional que solo muestra la tasa final, no las condiciones de mercado y, si acaso, alguna anotación manual en mi planilla de liquidez".

Tercero, la velocidad es una restricción crítica. En operaciones con clientes institucionales, los participantes describen ventanas de decisión muy cortas —en algunos casos, 15 minutos— en las que deben consultar, validar y cotizar una tasa. La consulta de múltiples fuentes sin una vista integrada introduce fricción operativa que puede traducirse en demoras o en el uso de referencias incompletas.

Cuarto, la descomposición de la tasa en componentes es una necesidad sentida. Múltiples informantes describieron de manera independiente la misma necesidad: una visualización que muestre de cuánto de la tasa final corresponde a la referencia de mercado, cuánto a la prima de liquidez, cuánto

al spread FTP y cuánto al ajuste por competencia. Esta descomposición, que es precisamente la lógica del basis de emisión implementada en el pipeline, fue mencionada espontáneamente por tres de los cinco participantes como la salida de mayor utilidad que podría ofrecer una herramienta analítica.

Hallazgos cuantitativos: frecuencias del proceso actual

Las preguntas de escala Likert sobre la frecuencia de ocurrencia de hechos en el proceso actual arrojaron los siguientes patrones:

Tabla 9

Frecuencias del Proceso Actual de Pricing Según Informantes Clave

Hecho evaluado	Nunca	Rara vez	A veces	Frecuente	Siempre
Se consulta una referencia de mercado comparable antes de cotizar	0	0	0	0	5
La competencia observada influye en la decisión final	0	0	0	0	5
La justificación de la tasa queda documentada de forma verificable	0	2	2	0	1
Se requiere escalar la decisión por falta de información integrada	0	0	2	0	3
La decisión depende del criterio de personas específicas	0	0	0	0	5

Nota. Respuestas de 5 informantes clave. Elaboración propia.

El patrón es claro: la consulta de referencias de mercado y la influencia de la competencia son prácticas universales en el grupo (5/5 con respuesta "Siempre"), lo que confirma que la información de mercado no está ausente. Sin embargo, la documentación verificable de la justificación es la dimensión más débil (solo 1 de 5 con respuesta "Siempre"; 2 con "Rara vez"), y la dependencia de criterio experto

individual es máxima (5/5 con "Siempre"). Esto confirma cuantitativamente el diagnóstico cualitativo: el problema no es la falta de datos, sino la ausencia de una capa integrada que convierta esos datos en una decisión trazable.

Uso de fuentes de información

Respecto al uso de fuentes específicas, todos los participantes reportaron una frecuencia "Muy alta" en el uso de curvas TES, curvas swap, referencias AAA, señales BanRep/macro, cotizaciones de competencia, señales de liquidez interna y, en menor medida, el histórico de decisiones propias. La unanimidad en este punto refuerza la validez del diseño del feature engineering del pipeline: las variables que los expertos usan en la práctica son exactamente las que el motor analítico integra como variables explicativas.

Criterios de adopción de la herramienta

Respecto a los criterios mínimos para confiar en el motor prescriptivo, los informantes identificaron de manera convergente cinco condiciones: (1) que la herramienta explique el origen de cada componente de la tasa recomendada, no que entregue un número sin justificación ("no acepto cajas negras"); (2) que se actualice con baja latencia para reflejar cambios intradía en las curvas; (3) que permita sobrescribir la recomendación con registro del motivo (override trazable); (4) que guarde un historial de decisiones y condiciones de mercado; y (5) que funcione como apoyo al criterio del trader, no como sustituto. Estas cinco condiciones están directamente incorporadas en el plan de intervención formulado en el capítulo correspondiente.

Respecto a la utilidad potencial de los entregables, todos los participantes calificaron como "Muy alta" la utilidad de los cinco módulos propuestos: recomendación de tasa/spread por plazo e indexador, comparación frente a referencias de mercado y competencia, explicación de los drivers de la recomendación, simulación de escenarios y registro trazable de la decisión. En cuanto a la valoración

global del proyecto, los cinco informantes lo calificaron con la máxima puntuación ("Muy alta"), y todos expresaron disposición a participar en revisiones posteriores de hallazgos o validaciones del diagnóstico.

Metodología de analítica y ciencia de datos: CRISP-DM

Para el componente analítico del estudio se adoptó CRISP-DM (Cross-Industry Standard Process for Data Mining), metodología que organiza proyectos de minería de datos y aprendizaje automático en seis fases iterativas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Campbell et al., 2020). CRISP-DM resulta especialmente pertinente porque permite iterar entre la perspectiva del negocio y la técnica, algo fundamental cuando el objetivo no es solo predecir, sino apoyar una decisión sensible de tesorería con requerimientos de trazabilidad, explicación y adopción por parte de usuarios no técnicos (IBM, s/f).

Tabla 10

Aplicación de CRISP-DM al Pipeline Analítico del Proyecto

Fase	Propósito	Actividades principales	Entregable principal
1. Comprensión del negocio	Precisar el problema y los criterios de éxito	Definición del costo de fondeo, escenarios de tasa fija e indexados, objetivos del motor.	Problema de decisión y criterios de valor
2. Comprensión de los datos	Inventariar y evaluar calidad de fuentes	Revisión de SX, SC, SB, SwapCC, series BanRep y documentos institucionales.	Mapa de fuentes y reglas de lectura
3. Preparación de los datos	Construir capas analíticas reproducibles	Descarga, parseo, persistencia Parquet, maestro TES, staging CDT, basis y dataset maestro.	Dataset limpio y trazable
4. Modelado	Entrenar y comparar algoritmos	Pipelines, modelos lineales y basados en árboles, validación y selección.	Modelos y ranking de desempeño
5. Evaluación	Valorar desempeño técnico y económico	RMSE, MAE, R^2 , realidad vs. predicción, importancia de variables.	Reporte de evaluación

6. Despliegue	Traducir el resultado en uso de negocio	Motor prescriptivo, app interactiva y lineamientos de gobierno.	Artefacto de uso y hoja de control
---------------	---	---	------------------------------------

Nota. Elaboración propia con base en Chapman et al. (2000) e IBM (s. f.).

Metodología para formular el plan de mejora de la intervención

La formulación del plan de mejora parte de la síntesis entre el diagnóstico empresarial y la evidencia analítica. El DOFA estratégico identifica problemas, restricciones y oportunidades (Uribe Macías, 2021); los resultados del pipeline muestran qué variables y reglas de negocio tienen mayor incidencia sobre el pricing; y el instrumento a informantes clave aporta criterios de adopción, riesgos percibidos y requisitos mínimos de uso. Esta triangulación evita que el plan sea una lista de deseos y lo convierte en una hoja de ruta realizable (Project Management Institute, 2021).

La hoja de ruta se estructura por frentes: datos, modelos, gobierno, adopción por usuario de negocio y monitoreo. La gestión por fases y por entregables incrementales es coherente con las buenas prácticas de dirección de proyectos (Project Management Institute, 2021) y con los principios de gobierno de modelos establecidos en SR 11-7 y NIST AI RMF.

Gobierno de datos, ética y gobierno del modelo

El diseño metodológico incorpora criterios explícitos de gobierno de datos y de modelo. Desde la perspectiva de riesgo de modelo, SR 11-7 establece que las entidades financieras deben contar con documentación, validación, gobernanza y controles proporcionales al uso de los modelos, especialmente cuando estos influyen en decisiones financieras relevantes (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011). En este trabajo, esos principios se traducen en trazabilidad de los insumos, separación entre etapas, control de leakage, documentación de variables, evaluación reproducible y necesidad de supervisión humana sobre las recomendaciones del motor.

Desde la perspectiva de IA responsable, el NIST AI RMF propone gestionar los riesgos a lo largo del ciclo de vida de los sistemas de inteligencia artificial para promover su desarrollo y uso confiable y responsable (Tabassi, 2023). Aplicado al presente caso, ello implica reconocer riesgos de sesgo por cobertura incompleta de mercado, dependencia excesiva de periodos atípicos, uso irreflexivo de variables competitivas y sobreconfianza en la salida del modelo. Por ello, la solución debe entenderse como soporte a la decisión y no como sustituto del criterio profesional de tesorería o ALM.

En cuanto al instrumento de levantamiento, se garantizarán participación voluntaria, consentimiento informado, anonimización de respuestas y uso estrictamente académico-empresarial de la información. El cuestionario no recolecta datos sensibles personales y las respuestas se reportarán de manera agregada o codificada para evitar identificar de manera individual a los participantes.

Resumen de la relación objetivos–metodología

La siguiente matriz sintetiza la correspondencia entre los objetivos específicos, la metodología aplicada y los productos verificables del estudio. Con ello se evidencia que el capítulo sí cumple la exigencia de la plantilla institucional de mostrar las fases que permiten alcanzar cada objetivo y la metodología utilizada para el diagnóstico y para la propuesta de intervención.

Tabla 11

Relación Objetivos–Metodología

Objetivo específico	Metodología principal	Fuentes / instrumentos	Producto esperado
Revisar referentes teóricos, sectoriales y regulatorios	Revisión documental estructurada	Matriz bibliográfica y documentos institucionales	Marco teórico y criterios de variables
Diagnosticar el proceso actual y sus brechas	Porter, PESTEL, DOFA y levantamiento AS-IS	Matrices diagnósticas y trazabilidad	Diagnóstico externo e interno
Estructurar un pipeline reproducible de datos	CRISP-DM: comprensión y preparación de datos	Inventario de fuentes, bitácoras y scripts	Dataset maestro reproducible

Desarrollar y validar el motor prescriptivo	CRISP-DM: modelado, evaluación e interpretabilidad	Matriz de variables, métricas y reportes	Modelo seleccionado y explicación de drivers
Formular lineamientos de adopción y evaluación de utilidad	Priorización impacto–viabilidad y hoja de ruta	DOFA integrado, hallazgos técnicos y criterios de gobierno	Plan de intervención empresarial

Nota. Elaboración propia con base en los objetivos

Diagnóstico Organizacional

El presente capítulo desarrolla el diagnóstico organizacional del proyecto a partir de la lógica metodológica ya definida en el trabajo: diagnóstico externo mediante PESTEL, Cinco Fuerzas de Porter y DOFA estratégico, junto con diagnóstico interno del proceso actual de pricing de CDTs, levantamiento AS-IS y revisión de la trazabilidad del pipeline analítico. En consecuencia, los hallazgos que aquí se presentan constituyen la línea base diagnóstica con la cual posteriormente se triangularán las respuestas del instrumento.

Propósito y alcance del diagnóstico

El propósito del diagnóstico organizacional es establecer el punto de partida real de la intervención, es decir, identificar en qué condiciones competitivas, operativas, analíticas y de gobierno se encuentra hoy el proceso de definición de tasas y spreads de CDTs en Corficolombiana. El diagnóstico no busca demostrar todavía la superioridad del motor prescriptivo, sino precisar por qué la organización requiere una solución de esta naturaleza y cuáles brechas debe cerrar para adoptarla de manera efectiva.

En coherencia con el objetivo específico de diagnosticar el proceso actual de pricing de CDTs, el capítulo se concentra en tres preguntas: (i) qué presiones del entorno afectan hoy el costo de fondeo y la decisión de pricing, (ii) cómo se soporta actualmente la decisión dentro de la organización y (iii) cuáles brechas separan ese proceso de una práctica más analítica, trazable y consistente. Este enfoque es consistente con los lineamientos de trabajo dirigido de la Universidad EAN, según los cuales el diagnóstico debe conducir de forma directa al plan de mejora y a la ruta de implementación del proyecto.

Diagnóstico externo

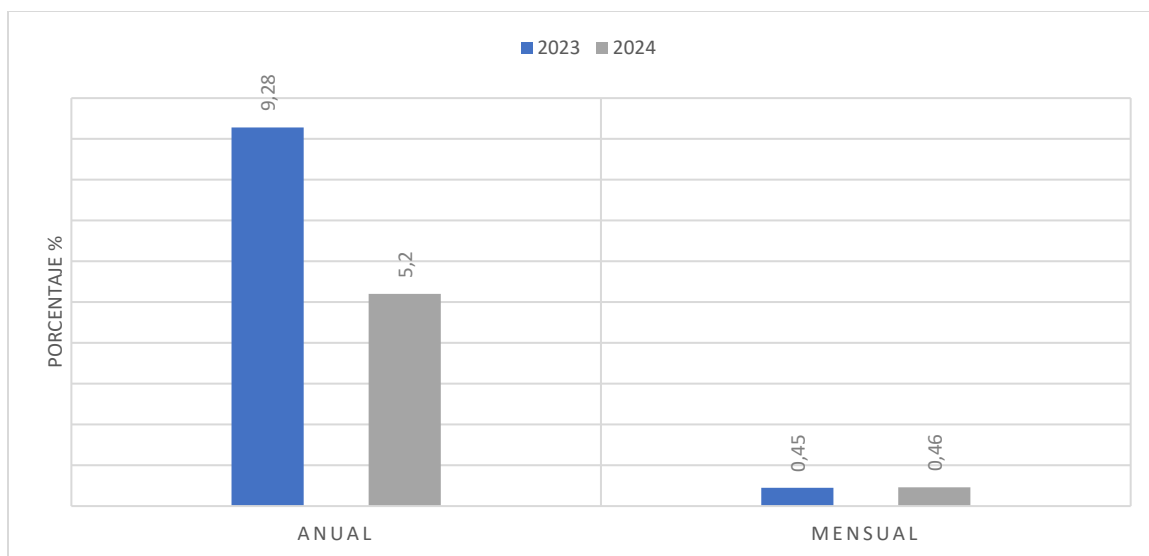
El diagnóstico externo sintetiza los hallazgos del análisis institucional y sectorial ya construidos en el trabajo. En lugar de repetir todo el marco institucional, esta sección traduce esa información en implicaciones directas para el problema de pricing. La lectura conjunta del entorno macroeconómico, la dinámica competitiva y las exigencias regulatorias permite concluir que Corficolombiana enfrenta un contexto en el que la definición de tasas de captación no puede depender exclusivamente de criterio experto o de comparaciones parciales, pues el costo de fondeo está altamente expuesto a cambios en tasas, liquidez y señales competitivas.

Tendencias del Sector y Entorno Macroeconómico

Durante 2024, la economía colombiana mostró una recuperación moderada. De acuerdo con el DANE, el Producto Interno Bruto presentó un crecimiento provisional de 1,5 % frente a 2023, impulsado principalmente por actividades como agricultura, administración pública y comercio (DANE, 2025a). En paralelo, la inflación anual a diciembre de 2024 se ubicó en 5,20 %, lo que reflejó una trayectoria descendente frente a los niveles observados en años previos y abrió espacio para una normalización gradual del entorno de tasas (DANE, 2025a). Este contexto macroeconómico es relevante para el problema de investigación porque las decisiones de pricing de captaciones están directamente influenciadas por la dinámica de inflación, política monetaria, liquidez y expectativas de mercado. La Figura 2 presenta esta evolución.

Figura 2

Índice de Precios al Consumidor

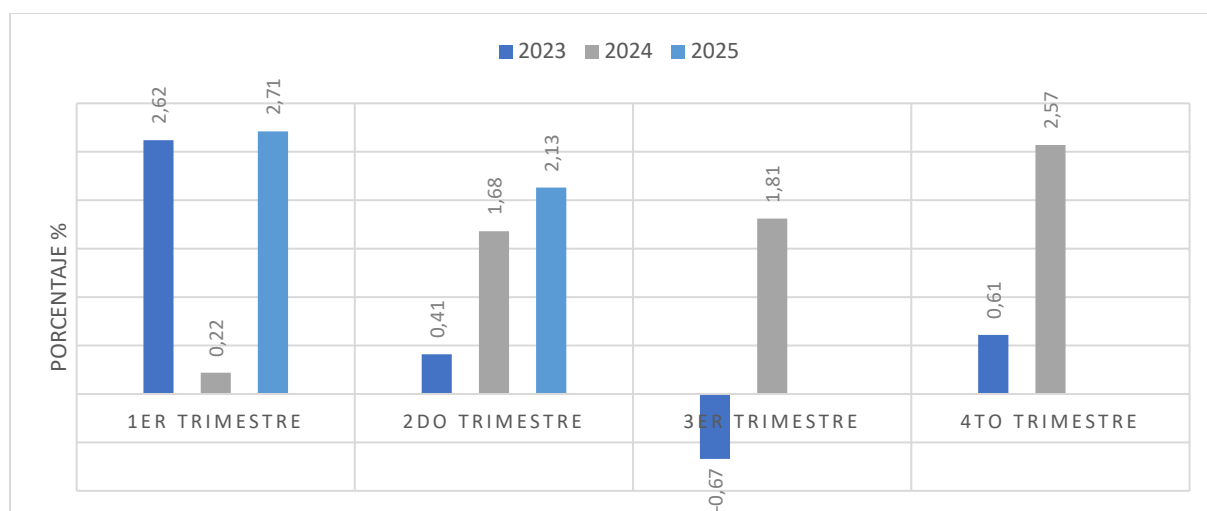


Nota. Adaptado de boletín técnico Índice de Precios al Consumidor (IPC), por DANE (2025a)

En 2025, la economía exhibió señales de recuperación. El boletín técnico del DANE para el primer trimestre de 2025 reporta que el Producto Interno Bruto creció 2,7 % en su serie original respecto al mismo trimestre de 2024 y 0,8 % frente al trimestre inmediatamente anterior (DANE, 2025b). La recuperación continuó en el segundo trimestre de 2025. La serie original del PIB creció 2,1 % frente al mismo periodo de 2024 y el acumulado del primer semestre se ubicó en 2,4 %, según el mismo organismo (DANE, 2025b), la Figura 3 resume la evolución trimestral del PIB. Estas cifras sugieren una aceleración gradual del crecimiento económico en 2025, aunque la inversión y algunos sectores siguen mostrando debilidad.

Figura 3

Crecimiento PIB Real, Trimestral Años 2023, 2024 y 2025



Nota. Adaptado de Producto Interno Bruto (PIB), por Banco de la República (2025)

Desde la perspectiva financiera, el Banco de la República concluyó en su Reporte de Estabilidad Financiera del segundo semestre de 2024 que los establecimientos de crédito mantienen niveles adecuados de capital y liquidez (Banco de la República, 2024a), pero también advirtió que las compras significativas de títulos de deuda pública aumentan la vulnerabilidad de las instituciones frente a movimientos adversos en la prima de riesgo soberano y en los mercados financieros locales (Banco de la República, 2024a). Esta observación resulta especialmente pertinente para Corficolombiana, dado su rol activo en renta fija, derivados y gestión de fondeo, y refuerza la necesidad de contar con herramientas de pricing que integren señales de mercado de manera estructurada y trazable.

A ello se suma una transformación estructural del sector financiero asociada con digitalización, mayor uso de datos y nuevos esquemas de interoperabilidad. En Colombia, la Superintendencia Financiera expidió la Circular Externa 004 de 2024, mediante la cual impartió instrucciones relativas a las finanzas abiertas y a la comercialización de tecnología e infraestructura a terceros (SFC, 2024). En el ámbito regional, el Banco Interamericano de Desarrollo ha señalado que las finanzas abiertas pueden fortalecer la competencia, ampliar el uso de productos financieros y elevar los estándares de trazabilidad y protección de datos (Herrera et al., 2023). En este contexto, el uso de ciencia de datos

para pricing de CDTs no debe verse como una herramienta aislada, sino como parte de una evolución más amplia hacia decisiones financieras basadas en información más integrada, comparable y verificable.

Competencia, información y formación de tasas

La presión competitiva también refuerza la necesidad de un esquema analítico más robusto. En un mercado donde los clientes institucionales y corporativos comparan referencias, plazos e indexadores, la decisión de pricing requiere homologar información de mercado y no solo reaccionar a tasas observadas de competidores. A esto se agrega la mayor disponibilidad de datos y el avance de ecosistemas de finanzas abiertas en la región, los cuales aumentan la velocidad con la que las señales de mercado se convierten en referencias comerciales y elevan el estándar de trazabilidad de las decisiones basadas en datos (Herrera et al., 2023).

Desde la óptica de las Cinco Fuerzas, el proyecto se ubica en un entorno de rivalidad alta, presión creciente de sustitutos y poder de negociación relevante por parte de clientes institucionales. Aunque las barreras regulatorias y de capital siguen siendo elevadas, la diferenciación sostenible no proviene solo del tamaño o de la calificación crediticia, sino también de la capacidad para convertir información dispersa en decisiones oportunas y consistentes. El diagnóstico externo sugiere entonces que la competitividad en pricing ya no depende únicamente de tener acceso a mercado, sino de procesar mejor las señales del mercado.

Tabla 12

Síntesis del Diagnóstico Externo

Factor	Hallazgo central	Implicación para el proyecto
Macroeconomía	Recuperación gradual del PIB, pero entorno aún sensible a tasas e inflación.	La decisión de pricing debe incorporar señales macro y monetarias de forma estructurada.

Estabilidad financiera	El sistema mantiene capital y liquidez adecuados, aunque con mayor vulnerabilidad al riesgo soberano.	Las tasas de captación no pueden separarse del contexto de mercado y de la curva soberana.
Competencia	Mayor presión por tasas y spreads en productos comparables.	Se requiere una metodología homogénea para comparar plazos, indexadores y referencias.
Tecnología y datos	Más disponibilidad de datos y mayor expectativa de decisiones trazables.	La solución debe ser reproducible, auditable y explicable.
Situación de Corfi	Tesorería con presión sobre margen y costo de fondeo material.	El problema es estratégico y económicamente relevante para la entidad.

Nota. Elaboración propia.

Diagnóstico interno del proceso actual de pricing de CDTs

El diagnóstico interno se construye a partir de cuatro fuentes complementarias: el proceso actual AS-IS, el inventario de fuentes disponibles, la arquitectura técnica desarrollada y los hallazgos del instrumento aplicado a cinco informantes clave durante el primer trimestre de 2026. Esta triangulación permite contrastar lo que el proceso hace en la práctica —según quienes lo ejecutan a diario— con lo que el pipeline revela sobre la estructura y dispersión de la información disponible. La conclusión de ambas lecturas es convergente: la organización no carece de datos ni de criterio experto, sino de una capa analítica integrada que conecte esos insumos con una recomendación de tasa o spread trazable y reproducible.

Situación AS-IS del proceso

La evidencia acumulada sugiere que el proceso actual de pricing parte de información valiosa y de conocimiento experto real, pero no de una arquitectura analítica plenamente institucionalizada. La organización cuenta con capacidad técnica y con acceso a fuentes robustas —Precia, series del Banco de la República, referencias de mercado y conocimiento de las mesas—; sin embargo, la integración de estos insumos en una única lógica decisional sigue siendo el principal cuello de botella. En otras palabras, la organización no carece de datos, sino de una capa suficientemente estandarizada para

convertir esos datos en una recomendación de tasa o spread consistente entre plazos, indexadores y contextos de mercado.

El proyecto mismo revela esa necesidad: fue necesario diseñar procesos de descarga automatizada, parsers específicos por tipo de archivo, capas de persistencia en Parquet, maestros de TES y de CDTs, curvas soberanas continuas, paneles de basis y datasets maestros para modelado. Ese esfuerzo no es accesorio; muestra que el problema organizacional no se reduce a elegir un algoritmo, sino a organizar un ecosistema de datos dispersos alrededor de una decisión concreta de tesorería.

Tabla 13

Levantamiento AS-IS del Proceso Actual de Pricing de CDTs

Etapa del proceso	Evidencia observada	Riesgo o limitación identificada
Consulta de mercado	Se utilizan curvas, valores de mercado, señales macro y referencias competitivas.	La información existe, pero puede consultarse de forma no homogénea entre actores o momentos.
Definición inicial de tasa/speed	La decisión incorpora lectura de mercado y criterio experto.	Alta dependencia de experiencia individual y de reglas tácitas.
Homologación por plazo e indexador	Se requieren comparaciones contra referentes soberanos, swap y competencia.	Sin una capa analítica integrada, las comparaciones pueden ser parciales o no estrictamente comparables.
Trazabilidad de la decisión	Hoy no toda la lógica de decisión queda necesariamente documentada en una sola ruta reproducible.	Dificultad para justificar ex post por qué se ofreció una tasa específica.
Aprendizaje y retroalimentación	La retroalimentación depende en gran parte del seguimiento operativo y de resultados posteriores.	Escasa institucionalización del aprendizaje histórico como insumo automático de decisiones futuras.

Nota. Elaboración propia.

Capacidades instaladas

El diagnóstico interno no es solo una lista de brechas. También muestra capacidades relevantes para que la intervención sea viable. Corficolombiana cuenta con una Vicepresidencia de Tesorería con mesas diferenciadas de ALM, moneda legal y moneda extranjera, así como una Vicepresidencia de

Gobierno, Riesgo y Cumplimiento y capacidades de tecnología y transformación digital en su organigrama corporativo (Corficolombiana, 2025c). Esta estructura es una fortaleza decisiva porque el proyecto requiere tanto conocimiento del negocio como condiciones mínimas de adopción y gobierno.

A esto se suma la disponibilidad de fuentes de información suficientemente ricas para construir el pipeline: archivos SX, SC, SB y SwapCC; series del Banco de la República; documentos institucionales; y artefactos técnicos ya desarrollados en Python. La existencia de estas fuentes demuestra que la organización tiene materia prima para operar con un motor prescriptivo. El problema no es la inexistencia de información, sino su dispersión, su heterogeneidad y la limitada institucionalización de la ruta analítica que las conecta con la decisión final.

Brechas identificadas

La principal brecha identificada es de integración analítica. Aunque la organización dispone de señales suficientes para alimentar la decisión de pricing, hoy esas señales no convergen necesariamente en una única estructura reproducible que permita explicar, comparar y recomendar tasas con criterios consistentes. Esta brecha se manifiesta en cuatro dimensiones: dependencia de criterio experto, trazabilidad incompleta, comparación imperfecta por plazo/indexador y uso no sistemático del aprendizaje histórico.

La segunda brecha es de formalización decisional. En el estado actual, la lógica de pricing puede ser financieramente razonable, pero no siempre queda encapsulada en un artefacto verificable que permita auditoría, seguimiento y comparación objetiva entre alternativas. Esto limita la gobernanza del proceso, sobre todo en un entorno financiero donde el uso de modelos y analítica avanzada exige documentación, validación y supervisión humana clara (Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency, 2011).

La tercera brecha es de adopción futura. Un motor prescriptivo solo genera valor si sus usuarios confían en la salida y pueden entender sus impulsores. Por ello, desde ya el diagnóstico deja ver que la solución no puede limitarse a un modelo con buen desempeño estadístico; debe incluir explicabilidad, reglas de uso, interfaces comprensibles y mecanismos de monitoreo. Esta brecha fue contrastada inicialmente mediante el instrumento aplicado a informantes clave y, posteriormente, mediante la validación funcional del prototipo con usuarios de mesa.

Tabla 14

Brechas Diagnósticas del Proceso Actual

Brecha	Descripción	Consecuencia organizacional
Integración de datos	Los insumos relevantes existen, pero no siempre están consolidados en una sola lógica decisional.	Mayor esfuerzo manual y menor consistencia entre decisiones.
Trazabilidad	La justificación completa de la decisión no siempre queda encapsulada en un flujo reproducible.	Dificultad para auditar, explicar o replicar decisiones.
Homologación de comparables	Plazos, indexadores y referentes de mercado requieren normalización técnica.	Riesgo de comparar instrumentos no estrictamente equivalentes.
Dependencia de criterio experto	El juicio experto es valioso, pero domina sobre una ruta analítica institucionalizada.	Vulnerabilidad a variaciones entre personas y momentos.
Gobierno del modelo futuro	La adopción exige controles, roles y explicación gerencial.	Riesgo de baja adopción o de uso acrítico de la salida del modelo.

Nota. Elaboración propia.

Análisis de las Cinco Fuerzas de Porter

Amenaza de Nuevos Entrantes.

Las barreras de entrada al sector financiero son altas en términos de capital y cumplimiento regulatorio; sin embargo, la digitalización y la normativa de open finance han reducido parte de estas barreras. La Circular Externa 004 de 2024 promueve la inclusión y competencia al permitir que

instituciones compartan datos financieros de clientes y desarrollen nuevos servicios (SFC, 2024). Esta apertura ha impulsado el crecimiento del ecosistema fintech; en 2024 operaban 394 fintech locales y 169 extranjeras en Colombia (FINNOSUMMIT, 2024). La entrada de neobancos y plataformas de cambio y crédito incrementa la presión competitiva sobre las mesas de dinero tradicionales.

Poder de Negociación de los Proveedores.

Los principales proveedores de Corficolombiana son fuentes de fondeo (bancos corresponsales, inversionistas institucionales) y proveedores tecnológicos. El poder de negociación es moderado: por un lado, la alta dependencia de financiamiento de corto plazo puede incrementar el costo de fondeo; por otro, la pertenencia al Grupo Aval y su sólida solvencia crediticia brindan acceso a recursos a costos competitivos. Las calificadoras destacan que el apoyo del conglomerado y la fortaleza patrimonial de Corficolombiana mitigan la presión de los proveedores y respaldan su capacidad para negociar condiciones favorables (Corficolombiana, s/f-c).

Poder de Negociación de los Clientes.

Los clientes corporativos, institucionales y gubernamentales tienen alternativas entre varios bancos universales y entidades financieras. El sector está concentrado: Bancolombia controla el 22,32 % de los activos del sistema, seguido de Davivienda (12,21 %), Banco de Bogotá (10,72 %), BBVA Colombia (8,88 %) y Banco de Occidente (6,16 %) (Rankings Latam, 2025). Esta concentración aumenta la competencia por clientes de alto valor, otorgándoles cierta capacidad de negociación. Sin embargo, la capacidad de Corficolombiana para estructurar operaciones complejas, su reputación en el mercado de deuda pública y su enfoque sostenible favorecen la fidelización de clientes clave.

Amenaza de Productos Sustitutos.

En las operaciones de tesorería, los sustitutos potenciales incluyen plataformas electrónicas de negociación, stablecoins, criptomonedas y servicios fintech que ofrecen tasas competitivas y acceso

inmediato. La regulación de open finance incentiva la creación de productos digitales y de inversión alternativos (Holland & Knight, 2024). Asimismo, el crecimiento del ecosistema fintech y el uso de inteligencia artificial en la negociación facilitan la aparición de soluciones que compiten con las mesas de dinero tradicionales(FINNOSUMMIT, 2024).

Rivalidad Entre Competidores.

La rivalidad en el sector financiero colombiano es intensa debido a la presencia de grandes bancos locales y multinacionales que compiten en tasas de interés, comisiones y servicios. El mercado está altamente concentrado: los cinco mayores bancos concentran más del 60 % de los activos y lideran las utilidades del sector (Rankings Latam, 2025). Corficolombiana compite diferenciándose por su conocimiento sectorial, su portafolio diversificado en infraestructura y energía, y su compromiso con la sostenibilidad, factores que le permiten mantener su posicionamiento a pesar de la fuerte rivalidad.

Tabla 15

Resumen Análisis de las Cinco Fuerzas de Porter

Fuerza	Análisis clave
Amenaza de nuevos entrantes	Aunque el capital regulatorio y las exigencias de solvencia son barreras altas, la digitalización y el open finance facilitan el ingreso de fintech y neobancos .
Poder de negociación de los proveedores	Los proveedores de fondeo (inversionistas, bancos corresponsales) ejercen presión al alza sobre las tasas, pero la calificación AAA de la entidad y el respaldo del Grupo Aval atenúan su poder.
Poder de negociación de los clientes	Los clientes cuentan con varias alternativas bancarias y fintech; su poder aumenta cuando las tasas de mercado suben y buscan mayores rendimientos.
Amenaza de productos sustitutos	Fondos de inversión, plataformas de préstamo colectivo y criptomonedas son sustitutos potenciales de los depósitos tradicionales, especialmente en un entorno de open finance.
Rivalidad entre competidores	La competencia es intensa debido a la presencia de grandes bancos, fintechs y neobancos; la entidad se diferencia por su experiencia en productos derivados y sostenibilidad.

Nota. Elaboración propia, resumen del apartado Análisis de las Cinco Fuerzas de Porter

Análisis PESTEL

Político.

El entorno político de Colombia se caracteriza por políticas orientadas al crecimiento y la sostenibilidad. La Superintendencia Financiera ha emitido la Circular Externa 004 de 2024 que crea el marco de open finance, fomentando la inclusión y competencia en el sector financiero (Holland & Knight, 2024). La estabilidad macroeconómica y la continuidad de los proyectos de infraestructura del Gobierno respaldan el desarrollo de sectores clave como energía y transporte.

Económico.

Tras un crecimiento del PIB del 0,7% en 2023 y de 1,5% en 2024 (DANE, 2025b), la economía colombiana muestra señales de recuperación. Sin embargo, el sector financiero enfrentó una contracción en la cartera de crédito, con una variación real anual de -5,2%, arrastrada por la caída del -10,7% en la cartera de consumo y -4,3 % en la cartera comercial, mientras que microcrédito (4,2%) y vivienda (1,3%) permanecieron en positivo (Findeter, 2024). La inflación anual se ubicó en 5,20 % en diciembre de 2024 (DANE, 2025a) y las tasas de interés continúan elevadas, lo que afecta el costo de fondeo y la rentabilidad de la tesorería.

Social.

La población colombiana demanda cada vez más servicios financieros digitales y productos de inversión sustentables. La creciente conciencia social sobre la sostenibilidad ha llevado a que Corficolombiana y otras instituciones integren criterios ESG en sus portafolios de inversión, como lo demuestra la meta corporativa de aumentar la participación de bonos verdes y sociales en sus carteras (Corficolombiana, 2025b). La aprobación de programas de inclusión financiera y la expansión del crédito a pequeñas y medianas empresas impulsan nuevas oportunidades para el sector.

Tecnológico.

La transformación digital es una tendencia clave en el sector financiero. Colombia cuenta con 394 fintech locales y 169 empresas extranjeras operando en 2024, mostrando un crecimiento del 6,8 % anual y posicionando al país como uno de los polos fintech de Latinoamérica (FINNOSUMMIT, 2024). El uso de algoritmos y tecnologías de inteligencia artificial en las mesas de dinero ha mejorado la eficiencia, la predicción estadística y la reducción de costos (Belltech, 2023). La regulación de open finance facilitará el intercambio de datos financieros y fomentará el desarrollo de nuevos productos digitales (Holland & Knight, 2024).

Ambiental.

La presión por la transición energética y el cumplimiento de los Objetivos de Desarrollo Sostenible impulsa proyectos de energías renovables. Corficolombiana participa en iniciativas de reforestación y proyectos agroindustriales sostenibles, lo que se alinea con su visión de generar valor económico, social y ambiental (Corficolombiana, s/f-c). La creciente demanda de financiamiento verde y bonos sostenibles abre oportunidades para ampliar la oferta de productos en la mesa de dinero y en la banca de inversión.

Legal.

El sector financiero colombiano está altamente regulado por la Superintendencia Financiera, el Banco de la República y la DIAN. La Circular Externa 004 de 2024 establece estándares tecnológicos, de ciberseguridad y protección de datos para el open finance (Holland & Knight, 2024). Adicionalmente, las políticas de riesgo, cumplimiento y gobierno corporativo imponen exigencias en términos de capital, revelación de información y prevención de lavado de activos y financiación del terrorismo.

Tabla 16

Resumen Análisis PESTEL

Factor	Principales hallazgos y evidencias
Político	Colombia impulsa la agenda de open finance mediante la Circular Externa 004 de 2024, fomentando la competencia y la apertura de datos; el entorno macroeconómico se mantiene estable.
Económico	La economía creció 1,5 % en 2024; sin embargo, la cartera de crédito se contrajo, la inflación se situó en torno al 5 % y las tasas de interés se mantuvieron elevadas, lo que incrementa el costo de fondeo.
Social	Existe una alta demanda de servicios financieros digitales y de productos sostenibles; los programas de inclusión financiera abren oportunidades.
Tecnológico	Se registraron 394 fintech locales y 169 extranjeras en 2024; la adopción de inteligencia artificial y la regulación de open finance están redefiniendo la banca.
Ambiental	La transición energética, los proyectos de reforestación y la demanda de financiamiento verde son clave; la entidad participa en la emisión de bonos verdes.
Legal	La regulación bancaria es exigente en materia de gestión de riesgos, transparencia y prevención de lavado de activos; la Circular 004 de 2024 refuerza los estándares de open finance y ciberseguridad.

Nota. Elaboración propia, resumen del apartado Análisis PESTEL.

Análisis DOFA (Fortalezas, Oportunidades, Debilidades y Amenazas)

Fortalezas.

Corficolombiana cuenta con una posición relevante en los mercados de renta fija, divisas y derivados, así como con una estructura organizacional que favorece una intervención analítica enfocada en tesorería. El informe CorfiSostenible 2024 muestra que la entidad mantiene una participación importante en el mercado spot peso-dólar, en forwards y en deuda pública, mientras que su organigrama evidencia capacidades diferenciadas en Tesorería, ALM, Riesgo, Cumplimiento y Tecnología (Corficolombiana, 2025b,2025c). A esto se suma la disponibilidad de información de mercado y de capacidades técnicas ya demostradas en el desarrollo del pipeline, lo que indica que la organización no parte de cero, sino de una base institucional y operativa sólida para soportar un motor prescriptivo de pricing.

Oportunidades.

El entorno actual ofrece condiciones favorables para fortalecer la toma de decisiones de pricing basada en datos. Por una parte, la agenda de finanzas abiertas (SFC, 2024) y comercialización de tecnología e infraestructura impulsada por la Superintendencia Financiera amplía las posibilidades de interoperabilidad, integración de información y desarrollo de soluciones analíticas más sofisticadas (SFC, 2024). Por otra, el BID ha señalado que las finanzas abiertas en América Latina y el Caribe pueden fortalecer la competencia, ampliar el uso de productos financieros y elevar los estándares de trazabilidad y protección de datos (Herrera et al., 2023). Además, la recuperación gradual de la economía y la corrección de inflación y tasas observada en 2025 pueden favorecer decisiones de captación más eficientes y mejor soportadas analíticamente (DANE, 2025b).

Debilidades.

La principal debilidad no está en la falta de información, sino en su fragmentación y en la ausencia de una lógica única e institucionalizada para la decisión de pricing de CDTs. Aunque la entidad dispone de curvas, referencias de mercado, series macroeconómicas y conocimiento experto, estos insumos no siempre convergen en una estructura reproducible que permita comparar de forma homogénea plazos, indexadores y alternativas de tasa. Esto se traduce en alta dependencia de criterio experto, trazabilidad parcial de las decisiones, dificultad para justificar ex post una cotización y uso limitado del aprendizaje histórico como insumo sistemático para decisiones futuras. En otras palabras, la mayor debilidad del proceso actual es la falta de una capa analítica integrada, explicable y gobernable.

Amenazas.

La principal amenaza proviene de un entorno en el que el costo de fondeo sigue siendo altamente sensible a la volatilidad de tasas, a la prima de riesgo soberano y a la competencia por recursos. El Banco de la República ha advertido que la mayor exposición del sistema financiero a títulos de deuda pública incrementa su vulnerabilidad frente a movimientos adversos en los mercados locales

(Banco de la República, 2024b). A esto se suma una mayor presión competitiva, una regulación más exigente en materia de datos y finanzas abiertas, y los riesgos asociados al uso de modelos e inteligencia artificial en finanzas, especialmente en términos de calidad de datos, gobernanza y control (Tabassi, 2023). Si el proceso de pricing no evoluciona hacia una lógica más analítica y gobernada, la organización corre el riesgo de seguir tomando decisiones heterogéneas, difíciles de auditar y menos eficientes frente al mercado.

Tabla 17
Resumen Análisis DOFA

Categoría	Aspectos destacados
Fortalezas	Posición relevante en tesorería y mercados, estructura apta para la adopción, disponibilidad de insumos y capacidades técnicas, respaldo institucional.
Oportunidades	Finanzas abiertas, mejor interoperabilidad de datos, mayor adopción de analítica en finanzas, entorno de tasas más favorable.
Debilidades	Fragmentación de datos, dependencia de criterio experto, trazabilidad parcial, comparabilidad imperfecta por plazo e indexador, falta de formalización del proceso de pricing.
Amenazas	Volatilidad de tasas y prima soberana, presión competitiva sobre captaciones, mayores exigencias regulatorias y de gobierno de datos y modelos.

Nota. Elaboración propia, resumen del apartado Análisis DOFA.

Priorización de problemas

No todas las brechas tienen la misma urgencia ni el mismo potencial de impacto. Para orientar el plan de intervención, se priorizaron los problemas según tres criterios: impacto sobre el costo de fondeo, efecto sobre la trazabilidad de la decisión y factibilidad de intervención mediante el pipeline desarrollado. Bajo estos criterios, el diagnóstico muestra que el problema más crítico no es la falta de información de mercado, sino la falta de una lógica integrada y prescriptiva para convertir esa información en decisiones repetibles.

Tabla 18*Priorización de Problemas del Diagnóstico*

Problema identificado	Impacto	Urgencia	Prioridad
Ausencia de una recomendación analítica única para pricing de CDTs	Alto	Alta	1
Trazabilidad parcial de la decisión comercial	Alto	Alta	2
Comparación no homogénea por plazo e indexador	Alto	Media	3
Dependencia elevada de criterio experto no formalizado	Media-Alta	Alta	4
Falta de lineamientos explícitos de adopción y monitoreo futuro	Media	Media	5

Nota. Elaboración propia.

Conclusión del diagnóstico

El diagnóstico organizacional permite concluir que Corficolombiana cuenta con condiciones institucionales, operativas y técnicas suficientes para adoptar una solución analítica de apoyo al pricing de CDTs, pero también evidencia brechas claras en integración de datos, trazabilidad de la decisión y formalización del proceso. En consecuencia, la necesidad de la intervención no surge de la falta de información de mercado, sino de la ausencia de una lógica integrada, reproducible y explicable que convierta esas señales en recomendaciones consistentes de tasa o spread.

En ese sentido, el problema identificado no es tecnológico en sentido estricto, sino organizacional y decisional. La entidad dispone de información valiosa, de experiencia de negocio y de capacidades técnicas, pero aún no ha consolidado una capa analítica que conecte de manera estructurada esos insumos con la decisión final de captación. Esta conclusión justifica plenamente la formulación del plan de intervención, cuyo propósito es institucionalizar esa capacidad dentro del proceso de tesorería bajo criterios de uso, control y monitoreo.

Desarrollo del pipeline analítico y resultados del modelado

Este capítulo documenta el desarrollo del pipeline reproducible de datos y los resultados del proceso de modelado y validación, dando respuesta directa a los objetivos específicos tres y cuatro del trabajo. El capítulo se organiza siguiendo las fases CRISP-DM adoptadas en el diseño metodológico: comprensión y preparación de los datos, construcción de variables, modelado, evaluación e interpretabilidad, y despliegue en una interfaz de usuario de negocio.

Los resultados técnicos reportados en este capítulo corresponden a la ejecución del pipeline realizada con corte de información al 03 de mayo de 2026. En consecuencia, los conteos de archivos procesados, registros, variables, métricas de desempeño, gráficos de interpretabilidad y salidas del prototipo reflejan el estado de las fuentes y artefactos disponibles hasta esa fecha. Las ejecuciones posteriores del pipeline pueden modificar de forma natural los volúmenes de datos, el ranking de modelos y las métricas de desempeño, por lo que la lectura de resultados se entiende acotada a ese corte analítico.

Arquitectura general del pipeline

El pipeline diseñado en este proyecto es una secuencia de trece componentes de software — organizados en diez scripts numerados más módulos de soporte— que transforman datos crudos de mercado en recomendaciones de pricing explicables para el usuario de tesorería. Cada componente tiene una entrada, una transformación definida y una salida persistida en formato Parquet, lo que garantiza la trazabilidad de cada dato desde su fuente original hasta la recomendación final. La Tabla 19 muestra el mapa completo del pipeline.

Tabla 19

Arquitectura del Pipeline Analítico por Componente

Capa	Script / componente	Insumo principal	Transformación principal	Salida / artefacto	Fase CRISP- DM
------	------------------------	---------------------	-----------------------------	-----------------------	-------------------

1. Ingesta macroeconómica	1.1_Banrep.py	API SDMX Banco de la República	Descarga, transformación y validación de series como IBR, TRM, UVR y tasa de política monetaria	Series macroeconómicas estructuradas y reporte de validación	Comprensión y preparación de datos
1. Ingesta Precia	1.2_Precia.py	Archivos SB, SC, SX y SwapCC	Descarga FTPS, organización por carpetas, manifest de auditoría y filtro inicial de instrumentos relevantes	Archivos crudos organizados y bitácora de descarga	Comprensión y preparación de datos
2. Maestro TES	2.1_Maestro_TESSX.py	Archivos SX con TES	Parsing posicional, extracción de TIR, precio, duración, convexidad, fechas y plazo	Maestro histórico de TES	Preparación de datos
2. Maestro CDT	2.2_Maestro_CDTSX.py	Archivos SX con CDTs	Identificación de CDTs, extracción de tasa, TIR, margen, calificación, emisor, fechas e ISIN	Maestro histórico de CDTs	Preparación de datos
3. Basis de emisión	3.1_Motor_Basis.py	Maestro CDT	Identificación de emisiones primarias, filtro de rezago Precia, selección de emisores comparables y cálculo de prima inicial	panel_basis_entrenamiento.parquet	Preparación de datos
4. Curva soberana PCHIP	3.2_Curva_TES_Pchip.py	Maestro TES	Interpolación PCHIP por fecha y plazo; construcción de curva fija y UVR; cálculo de pendiente 10Y-1Y	curva_tes_pchip.parquet	Preparación de datos
5. Curvas SC/SB/SwapCC	4.1_Maestro_CceroC.py	Archivos SC, SB y SwapCC	Procesamiento de curvas cero cupón, betas y curva swap; homologación de tasas	Parquets de curvas cero cupón, betas y swap	Preparación de datos
6. Feature engineering	4.2_Feature_Engineering_Macro.py	Panel basis, curvas, series macro y competencia	Cruces por fecha, plazo e indexador; construcción de variables soberanas, AAA, swap, macro y top bancos	dataset_ml_master.parquet	Preparación de datos
7. Preproces	5_Preprocesamiento_ML.py	Dataset maestro	Deduplicación, control de rangos, normalización, control	dataset_ml limpio.parquet	Preparación de datos

amiento ML			de leakage, anclas temporales y categorización		
8. Modelado supervisa do	6_Modelado _ML_Avanza do.py	Dataset limpio	Entrenamiento temporal por escenarios, búsqueda de hiperparámetros, comparación de modelos y selección de ganadores	Modelos .joblib, métricas CSV y metadatos JSON	Modelado y evaluación
9. Interpreta bilidad	7_Interpreta bilidad.py	Modelos ganadores y dataset limpio	Importancia por permutación, coeficientes agregados, importancia nativa, residuales, correlaciones y deciles	Gráficos, métricas y CSV de interpretabilidad	Evaluación
10. Motor prescripti vo / app	8_App_Strea mlit.py	Dataset limpio, modelos y artefactos	Interfaz de consulta, torneo de modelos, drivers, simulación de tasa/spread y curva completa	Prototipo Streamlit de apoyo a tesorería	Despliegue

Nota. Elaboración propia a partir de los scripts desarrollados en el proyecto.

Capa 1. Ingesta de información macroeconómica y de mercado

La primera capa del pipeline se orientó a la adquisición automatizada de información externa. Para las variables macroeconómicas y monetarias se desarrolló el script 1.1_Banrep.py, el cual implementa un cliente de consulta a la API SDMX del Banco de la República. Este cliente extrae series históricas como IBR, TRM, UVR y tasa de política monetaria, transforma la información desde XML a estructuras tabulares de pandas y genera un archivo Excel con hojas y gráficos de validación.

Estas series cumplen una función contextual dentro del modelo. El IBR actúa como referencia monetaria para instrumentos indexados; la UVR permite aproximar condiciones asociadas a inflación; la TRM captura presiones cambiarias que pueden afectar expectativas de mercado; y la tasa de política monetaria refleja la postura del Banco de la República. La generación del archivo Excel no se utilizó únicamente como reporte operativo, sino también como mecanismo de validación visual para identificar

discontinuidades, valores atípicos o posibles fallas de extracción antes de integrar las series al dataset maestro.

De forma complementaria, el script 1.2_Precia.py automatizó la descarga de archivos de mercado desde Precia mediante conexión FTPS segura. Este proceso descarga archivos SB, SC, SX y SwapCC para un rango de fechas definido, organizándolos en carpetas locales diferenciadas y registrando cada ejecución en un manifest de auditoría. Dicho manifest incluye fecha, tipo de archivo, ruta remota, archivo descargado, ruta local, estado del proceso y detalle de eventuales errores.

En esta capa se aplicó un primer criterio de depuración sobre los archivos SX. Dado que estos archivos contienen un volumen amplio de instrumentos de renta fija, el script conservó los registros asociados a TES —identificados mediante prefijos como TFIT, TUVT y TFVT— y títulos con calificaciones de alta calidad crediticia, específicamente AAA, F1+ y BRC1+. Esta regla permitió reducir el volumen de procesamiento sin perder los instrumentos relevantes para construir referencias soberanas, referencias privadas de alta calidad y observaciones comparables de captación.

La decisión de filtrar por calificaciones altas responde a un criterio metodológico de comparabilidad. El pricing de un CDT no depende únicamente del plazo o del indexador, sino también del riesgo crediticio percibido del emisor. Por tanto, incluir emisores o instrumentos con calificaciones sustancialmente inferiores podría introducir diferenciales de tasa explicados por riesgo de crédito y no por condiciones de liquidez, competencia o estructura de tasas. Al concentrar el análisis en instrumentos de alta calidad crediticia, el pipeline reduce ruido y mejora la consistencia del universo de modelado.

Capa 2. Construcción de maestros de TES y CDTs

La segunda capa transformó los archivos SX descargados en maestros estructurados de instrumentos financieros. El script 2.1_Maestro_TESSX.py reconstruyó el maestro histórico de TES, extrayendo títulos soberanos con prefijos como TFIT, TUVT y TFVT. Para cada instrumento se capturaron

variables como fecha de valoración, fecha de emisión, vencimiento, moneda, tasa cupón, precio sucio, precio limpio, TIR, duración, duración modificada y convexidad.

Un aspecto técnico relevante es que el script implementó un parser con offset dinámico para leer formatos antiguos y nuevos de archivos SX. Esta decisión fue necesaria porque la estructura posicional de estos archivos no se mantiene completamente constante en el tiempo. El parser identifica la posición del marcador del registro y ajusta las posiciones de lectura según el formato detectado. Posteriormente calcula el plazo de cada TES bajo una convención ACT/365 ajustada por años bisiestos, lo que permite representar cada referencia soberana en una escala homogénea de plazo.

Este maestro de TES constituye la base para la construcción de curvas soberanas interpoladas. En otras palabras, el pipeline no depende exclusivamente de nodos de curva previamente publicados, sino que reconstruye una base histórica de instrumentos soberanos observados para generar referencias por fecha y plazo.

En paralelo, el script `2.2_Maestro_CDTsX.py` construyó el maestro histórico de CDTs. Este proceso identifica registros de CDTs mediante reglas sobre nemotécnico e ISIN, extrae variables financieras y operativas, y enriquece cada registro con el nombre del emisor utilizando un maestro auxiliar de NEMOS e ISIN.

Este maestro es fundamental porque contiene las condiciones observadas de los CDTs del mercado: emisor, fecha de emisión, fecha de vencimiento, tipo de tasa, tasa facial, TIR, margen, calificación y plazo. Además, conserva la trazabilidad del archivo de origen y la fecha de archivo, lo cual permite reconstruir la información utilizada para cada observación.

Antes del procesamiento masivo, el script genera un preview de validación en Excel y CSV. Esta etapa permite revisar una muestra de registros antes de procesar toda la historia, verificando que variables como tasa, TIR, margen, calificación, tipo de tasa y emisor estén siendo interpretadas

correctamente. En la ejecución reportada, el ETL de CDTs procesó 4 archivos nuevos, omitió 2.667 archivos ya existentes, identificó 9 archivos vacíos y no presentó errores. Esto evidencia una lógica incremental estable y controles básicos de calidad sobre el procesamiento.

Figura 4

Resultados Pipeline 2.2_Maestro_CDT SX.py

```
=====
RESUMEN ETL SX CDT
=====
Procesados: 4
Saltados: 2667
Vacíos: 9
Errores: 0
=====
```

Nota. Elaboración propia a partir de la ejecución del script.

Capa 3. Cálculo de la prima o basis de emisión

La tercera capa construyó el panel base de modelado mediante el script 3.1_Motor_Basis.py. Su objetivo fue identificar emisiones primarias de CDTs y calcular una prima inicial de emisión, denominada Prima_Emision_Basis, que sirve como variable central para el análisis supervisado.

En esta etapa se incorporó el criterio de selección de entidades comparables. El script no se limitó a Corficolombiana, sino que incluyó un grupo de emisores financieros relevantes del mercado colombiano: Corficolombiana, Banco de Bogotá, Davivienda, BBVA Colombia, Bancolombia, Banco de Occidente, Banco Agrario, Banco Popular, Findeter, AV Villas, Banco Caja Social, Scotiabank Colpatría e Itaú. Este grupo fue parametrizado en el código mediante la lista de entidades comparables y utilizado para filtrar el maestro de CDTs.

La selección de estas entidades no busca representar la totalidad del mercado financiero colombiano, sino construir un universo competitivo pertinente para decisiones de pricing de CDTs. Se priorizaron emisores con presencia recurrente en los archivos históricos, relevancia en captación institucional, disponibilidad suficiente de información y condiciones de riesgo más comparables con una

entidad financiera vigilada. La inclusión de Corficolombiana permite que el modelo aprenda tanto del comportamiento histórico propio como del comportamiento observado en pares de mercado.

La identificación de emisiones primarias se realizó tomando la primera observación disponible por ISIN. Este criterio permite aproximar el momento en que el instrumento aparece por primera vez en la información de valoración de Precia. Luego se calculó la variable `Días_Rezago_Precia`, definida como la diferencia entre la fecha del archivo de Precia y la fecha oficial de emisión del CDT. Solo se conservaron emisiones con rezago entre 0 y 8 días. Este control es importante porque evita tratar valoraciones secundarias tardías como si fueran condiciones cercanas a la emisión original.

Posteriormente se homologó el tipo de tasa. Las referencias FS y F se clasificaron como tasa fija, mientras que IB1 se clasificó como IBR. Esta homologación reduce la dispersión de códigos operativos y permite comparar instrumentos por familias de indexación.

El cálculo de la prima o basis inicial se realizó de manera diferenciada. Para CDTs de tasa fija, la prima se calculó como la diferencia entre la TIR observada por Precia y la tasa facial del título. Para instrumentos indexados, se calculó como la diferencia entre el margen observado y la tasa o spread registrado. Esta variable no debe confundirse con el spread completo frente a la curva soberana o frente a una curva AAA; esa comparación se enriquece posteriormente en el feature engineering. En esta capa, el objetivo es capturar la prima inicial observada en la emisión primaria.

En la ejecución reportada, el script procesó 2.671 archivos históricos de CDTs, extrajo 85.089.267 valoraciones diarias del grupo de entidades seleccionadas y validó 431.191 emisiones primarias únicas. Los cinco emisores con mayor presencia en el panel fueron Banco de Bogotá, BBVA Colombia, Banco Popular, Bancolombia y Banco de Occidente. El resultado final se guardó en `panel_basis_entrenamiento.parquet`.

Esta capa es especialmente relevante porque define la base supervisada del proyecto. El modelo no se entrena sobre observaciones arbitrarias del mercado secundario, sino sobre emisiones primarias depuradas, con control de rezago, homologación de indexador y delimitación de emisores comparables.

Figura 5

Resultados Pipeline 3.1_Motor_Basis.py

```
=====
● EXTRACTOR DE BASIS ECO-CÉNTRICO (TOP MERCADO COLOMBIANO)
=====
Extrayendo Top Bancos: 100%|
[~] Total de valoraciones diarias (históricas) del Top Mercado: 85,089,267
[~] Emisiones primarias únicas (Día 1) validadas: 431,191

=====
RESUMEN EXTRACTOR DE BASIS MULTI-BANCO
Emisor
BCO. DE BOGOTA      134377
BBVA COLOMBIA S.A   121662
BCO. POPULAR        105766
BCO. BANCOLOMBIA    17163
BCO. DE OCCIDENTE   15912
=====
Panel ML guardado en: panel_basis_entrenamiento.parquet
=====
```

Nota. Elaboración propia a partir de la ejecución del script.

Capa 4. Construcción de curvas soberanas interpoladas mediante PCHIP

La cuarta capa construyó curvas soberanas continuas mediante el script

3.2_Curva_TES_Pchip.py. Su propósito fue generar una tasa de referencia para cualquier plazo en días, de forma que cada CDT pudiera compararse contra una curva soberana equivalente en fecha e indexador.

La necesidad de esta etapa surge porque los TES observados en el mercado no existen para todos los plazos posibles. Un CDT puede emitirse a 90, 180, 360, 540, 720 o más días, mientras que las referencias soberanas tienen vencimientos específicos. Para resolver esta diferencia, se utilizó interpolación PCHIP, una técnica cúbica por tramos que preserva la forma de la curva y evita oscilaciones artificiales entre nodos observados.

El script filtra el maestro de TES por tipo de instrumento. Para la curva de tasa fija utiliza TES TFIT; para la curva real UVR utiliza TES TUVT; y para una eventual curva IBR intenta utilizar referencias TIBR. En la ejecución reportada se construyeron exitosamente la curva de tasa fija y la curva real UVR, mientras que no se encontraron datos suficientes para el prefijo TIBR.

La metodología exige al menos tres puntos observados por fecha para construir una curva no lineal. Además, elimina plazos duplicados y controla extrapolaciones en los extremos, manteniendo plano el valor de la tasa más cercana cuando el plazo objetivo se encuentra fuera del rango observado. Estas reglas reducen el riesgo de generar tasas espurias en segmentos con baja disponibilidad de referencias.

El archivo resultante se guardó como `curva_tes_pchip.parquet` e incluyó las variables Fecha, Plazo_Dias, TIR_TES_PCHIP, TIR_TES_UVR_PCHIP y Pendiente_Fija_10Y_1Y. Esta última mide la diferencia entre la tasa soberana fija a 10 años y la tasa a 1 año, capturando la forma de la curva. En términos financieros, esta pendiente permite incorporar información sobre empinamiento, aplanamiento y expectativas de mercado por plazo.

Esta capa es crítica porque convierte la información soberana observada en una referencia granular y comparable para cada emisión de CDT. Sin esta interpolación, el modelo tendría que comparar instrumentos con plazos no equivalentes, lo que afectaría la precisión del cálculo de spreads.

Figura 6

Resultados Pipeline 3.2_Curva_TES_Pchip.py

```
=====
co GENERADOR DE CURVAS SOBERANAS MULTI-INDEXADOR (MÉTODO PCHIP)
=====
[*] Cargando maestro de mercado TES...
[2/3] Construyendo Curva Tasa Real UVR (TU)...
[3/3] Construyendo Curva IBR (TIBR)...
[!] No se encontraron datos para el prefijo TIBR
[*] Consolidando curvas...
[*] Extrayendo Pendientes (Slope) 10Y-1Y de las curvas...

[OK] Mega-Curva PCHIP guardada exitosamente.
-> D:\Precia\Precia\processed3\curvas_interpoladas\curva_tes_pchip.parquet

Columnas Generadas:
['Fecha', 'Plazo_Dias', 'TIR_TES_PCHIP', 'TIR_TES_UVR_PCHIP', 'Pendiente_Fija_10Y_1Y']
```

Nota. Elaboración propia a partir de la ejecución del script.

Capa 5. Procesamiento de curvas cero cupón, curvas swap y betas

La quinta capa procesó los archivos SC, SB y SwapCC de Precia mediante el script

4.1_Maestro_CceroC.py. Esta etapa convirtió archivos crudos de curvas cero cupón, parámetros de curvas y curvas swap en Parquets estructurados y reutilizables.

Para los archivos SC, el script leyó estructuras de ancho fijo y extrajo campos como fecha, curva, plazo en días, tasa continua y tasa efectiva anual. Estas curvas permiten incorporar referencias como la curva soberana CEC y curvas privadas AAA por modalidad. Para los archivos SB, el script extrajo parámetros de curvas como Beta 0, Beta 1, Beta 2 y Beta 3, que capturan componentes de nivel, pendiente y curvatura de la estructura temporal. Para los archivos SwapCC, el script leyó días y tasas, limpió valores no numéricos y transformó las tasas a una tasa efectiva anual base 365.

Esta homologación de tasas es relevante porque las fuentes originales pueden expresarse bajo convenciones diferentes. Al llevarlas a una base comparable, el pipeline reduce inconsistencias y permite cruzar curvas soberanas, curvas privadas y curvas swap dentro de una misma matriz analítica.

En la ejecución reportada, el proceso escaneó 8.039 archivos en total, de los cuales 8.027 ya estaban procesados y fueron conservados como histórico intacto. Se procesaron 12 archivos nuevos en formato Parquet y no se presentaron errores. Por carpeta, se identificaron 2.679 archivos válidos de SwapCC, 2.680 archivos válidos de SC y 4 archivos nuevos de SB. Este resultado evidencia que el data lake se mantuvo actualizado sin reprocesar innecesariamente todo el histórico.

Figura 7

Resultados Pipeline 4.1_Maestro_CceroC.py

```
=====
INICIANDO ETL DATA LAKE (PRECIA)
Origen: D:\Precia\Precia
Destino: D:\Precia\Precia\processed2
=====

Escaneando carpeta: [SwapCC] ...
├ Total válidos en carpeta: 2679
├ Ya existen (saltados): 2675
├ Nuevos por procesar: 4
└ Completado: 4 archivos guardados | ✖ Errores: 0

Escaneando carpeta: [SC] ...
├ Total válidos en carpeta: 2680
├ Ya existen (saltados): 2676
├ Nuevos por procesar: 4
└ Completado: 4 archivos guardados | ✖ Errores: 0

Escaneando carpeta: [SB] ...
├ Nuevos por procesar: 4
└ Completado: 4 archivos guardados | ✖ Errores: 0

=====
RESUMEN EJECUTIVO ETL
=====
Total Archivos Escaneados: 8039
Histórico Intacto (Saltos): 8027
NUEVOS Procesados y Parquet: 12
=====
```

Nota. Elaboración propia a partir de la ejecución del script.

Esta capa aporta insumos esenciales para el feature engineering: curvas soberanas, curvas privadas AAA, curvas swap y parámetros de forma de curva. Estas referencias permiten que el modelo

no dependa únicamente de la tasa observada de cada CDT, sino que incorpore señales de mercado comparables por fecha y plazo.

Capa 6. Feature engineering macrofinanciero y construcción del dataset maestro

La sexta capa integró todos los insumos anteriores mediante el script 4.2_Feature_Engineering_Macro.py. Esta etapa es el núcleo de la preparación de datos para machine learning, porque transforma el panel de emisiones primarias en un dataset enriquecido con variables de mercado, curvas, competencia, liquidez y señales macroeconómicas.

El proceso inició con un panel base de 431.191 CDTs emitidos en 2.545 días distintos. Luego se cargó la curva soberana PCHIP y se generó una auditoría previa sobre 9 fechas representativas. Esta auditoría permitió validar que los cruces por fecha, plazo e indexador funcionaran correctamente antes de ejecutar el procesamiento masivo. El resultado de este control fue el archivo AUDITORIA_PREVIA_FEATURE_ENG.xlsx.

El cruce de fuentes se realizó utilizando llaves financieras y no únicamente fechas. Para las curvas se usaron combinaciones de fecha de emisión y plazo original en días. Para las señales competitivas se utilizaron fecha, tipo de tasa homologado y plazo. Esta decisión es importante porque el pricing de CDTs depende simultáneamente del momento de mercado, el vencimiento del instrumento y el indexador utilizado.

En esta capa también se incorporó el grupo de entidades comparables para construir señales competitivas. El script calcula indicadores agregados de tasa, TIR y margen promedio por fecha, tipo de tasa y plazo. Además, genera variables individuales por emisor mediante tablas pivote. De esta forma, el dataset final captura dos niveles de competencia: una señal consolidada del grupo comparable y señales particulares de bancos relevantes.

Entre las variables más importantes generadas por el feature engineering se encuentran:

Tabla 20*Variables Feature Engineering*

Variable	Interpretación
Tasa_Modelo_AAA_Exacta	Curva AAA aplicable según el tipo de indexador del CDT.
Curva_Soberana_Base	Referencia soberana aplicable al instrumento según su indexador y plazo.
Spread_Banco_vs_Soberano	Diferencia entre la TIR del CDT y la curva soberana comparable.
Tasa_Facial_TopBancos	Promedio de tasa facial observada en el grupo de bancos comparables.
TIR_Real_TopBancos	Promedio de TIR observada en el grupo comparable.
Margen_Real_TopBancos	Promedio de margen para instrumentos indexados del grupo comparable.
Premium_Pagado_Sobre_TopBancos	Diferencia entre la condición del CDT y el promedio comparable del mercado.
Spread_Vitrina_vs_TopBancos	Diferencia entre la tasa facial propia y la tasa facial promedio del grupo comparable.
Estres_Liquidez_Spot_Swap	Diferencia entre la curva swap y el IBR a tres meses, como señal de presión de liquidez o mercado.

Nota. Elaboración propia a partir de la ejecución del script.

Estas variables traducen conocimiento financiero en insumos modelables. Por ejemplo, Spread_Banco_vs_Soberano permite saber cuánto paga una emisión frente a una referencia soberana comparable; Premium_Pagado_Sobre_TopBancos mide si la entidad está pagando más o menos que sus pares; y Estres_Liquidez_Spot_Swap incorpora una señal de tensión entre el mercado monetario spot y la curva de derivados.

Dado el volumen de información, el procesamiento masivo se realizó por lotes o chunks. En la ejecución reportada se procesaron 9 lotes, lo que permitió fusionar fuentes de gran tamaño sin saturar memoria. El resultado final fue un dataset maestro de 1.673.550 registros y 88 columnas, guardado como dataset_ml_master.parquet.

Esta base constituye la matriz analítica principal para las etapas posteriores de limpieza final, entrenamiento, evaluación, selección del modelo e interpretabilidad. Desde el punto de vista metodológico, esta capa materializa la fase de preparación de datos de CRISP-DM, ya que convierte fuentes dispersas en una tabla de aprendizaje supervisado con variables financieras relevantes y trazables.

Figura 8

Resultados Pipeline 4.2_Feature_Engineering_Macro.py

```
=====
MOTOR DE FEATURE ENGINEERING (CRUCE VECTORIZADO OPTIMIZADO)
=====
-> Panel Base: 431,191 CDTs emitidos en 2545 días distintos.
[~] Cargando Curva Soberana Real (PCHIP)...

*****
GENERANDO AUDITORÍA PREVIA PARA 9 FECHAS
*****
[~] Cargando Curvas SC en RAM...
Curvas SC: 100%|
[~] Cargando Curvas Swap...
Curvas Swap: 100%|
[~] Calculando Índice de Competencia Individual y Consolidado...
Escaneando SX: 100%|
[OK] Auditoría guardada en: AUDITORIA_PREVIA_FEATURE_ENG.xlsx

[?] ¿Deseas continuar con el cruce masivo de toda la historia? (s/n): s

[+] Aprobado. Iniciando carga de matrices base...
[~] Cargando Curvas SC en RAM...
Curvas SC: 100%|
[~] Cargando Curvas Swap...
Curvas Swap: 100%|
[~] Calculando Índice de Competencia Individual y Consolidado...
Escaneando SX: 100%|
[~] Fusionando de forma particionada (Chunking)...
Procesando Lotes: 100%|
[~] Consolidando dataset final...

-----
FEATURE ENGINEERING COMPLETADO CON ÉXITO
-----

[+] Registros Enriquecidos y Listos para ML: 1,673,550
[+] Columnas procesadas: 88
[+] Dataset Final guardado en: dataset_ml_master.parquet
-----
```

Nota. Elaboración propia a partir de la ejecución del script.

Capa 7. Preprocesamiento final para machine learning y control de leakage

Después de construir el dataset maestro enriquecido, se desarrolló una etapa de preprocesamiento final mediante el script 5_Preprocesamiento_ML.py. Esta capa tuvo como propósito convertir el archivo dataset_ml_master.parquet en una base limpia, deduplicada, temporalmente ordenada y apta para entrenamiento supervisado, almacenada posteriormente como dataset_ml_limpio.parquet.

Esta etapa es importante porque el dataset maestro contiene información de múltiples fuentes, variables auxiliares de trazabilidad, fechas, identificadores, campos operativos y variables derivadas. No todos estos campos deben ingresar al modelo. Algunos son necesarios para auditar el pipeline, pero no son variables que el modelo deba usar para aprender; otros pueden introducir ruido o riesgo de fuga de información. Por ello, el preprocesamiento final funciona como una capa de control entre la ingeniería de características y el modelado.

El primer paso del script consiste en normalizar columnas base. Para ello se homologan nombres de fechas, se asegura la existencia de fecha_emision, fecha_vencimiento y plazo_original_dias, y se estandariza el tipo de tasa mediante la variable tipo_tasa_ref_homologado. En esta homologación, referencias como FS, F o FIJA se agrupan como tasa fija; IB1 e IBR se agrupan como IBR; y IPC se conserva como indexador propio. Esta normalización reduce la dispersión de códigos operativos y permite que los modelos trabajen con familias de instrumentos comparables.

Luego se construye la variable fecha_mercado, que representa la mejor aproximación disponible a la fecha de observación de mercado. Para construirla, el script utiliza una jerarquía de fechas posibles: Fecha Registro, fecha_corte_precia, fecha_emision o Fecha de Emisión. Esta decisión permite preservar una referencia temporal incluso cuando algunas fuentes tienen nombres distintos o campos incompletos.

Posteriormente, el script normaliza la escala de variables numéricas como `Prima_Emision_Basis`, Tasa, TIR y Margen. Cuando detecta que la mediana de una variable es mayor a 1, la divide entre 100 para expresarla en formato decimal. Esta regla evita mezclar tasas en porcentaje con tasas en decimal, un problema frecuente en bases financieras integradas desde distintas fuentes.

Una de las transformaciones más relevantes es la deduplicación temporal. El script elimina duplicados utilizando identificadores como ISIN, nemotécnico o número de emisión, combinados con la fecha de mercado. Esta deduplicación busca conservar una única observación válida por instrumento y fecha, evitando que un mismo título aparezca repetido y sesgue el entrenamiento. En la ejecución reportada, la base pasó de 1.673.550 registros brutos a 431.191 registros tras deduplicación temporal.

El script también aplica filtros de rango para remover valores extremos o financieramente inconsistentes. En particular, conserva la `Prima_Emision_Basis` dentro de un rango razonable entre -15 % y 15 %, y mantiene tasas y TIR entre 0 % y 35 %. Esta regla no busca eliminar volatilidad real del mercado, sino excluir registros que puedan corresponder a errores de lectura, problemas de escala o valores atípicos no útiles para entrenamiento.

Después de la limpieza, se construyen variables adicionales para modelado. Para instrumentos indexados a IBR o IPC, se determina el margen como `Spread_Indexado`. También se crean agrupaciones de plazo: `Plazo_Agrupado`, con categorías como corto, medio, largo y muy largo; y `Bucket_Ancla`, con tramos como 0-180, 181-360, 361-720, 721-1080 y >1080. Estas categorías permiten que el modelo capture patrones diferenciados por vencimiento.

Una innovación relevante de esta capa es la construcción de anclas temporales. El script calcula medianas móviles rezagadas para tasa fija y spreads indexados, agrupando por tipo de tasa, bucket de plazo y calificación. Estas anclas se calculan con `shift(1)`, es decir, usando únicamente información anterior a la observación actual. Este punto es metodológicamente importante porque reduce el riesgo

de leakage: el modelo no recibe información del mismo registro ni de observaciones futuras para explicar el target. A partir de estas anclas se generan variables como Ancla_Tasa_Fija, Ancla_Spread_Indexado, Residual_Tasa_Fija y Residual_Spread.

Finalmente, el script elimina columnas consideradas ruido o riesgo de leakage, como Secuencia, Tipo, Archivo_Origen, Dias_Rezago_Precia y fecha_corte_precia. Estas variables fueron útiles para construir y auditar la base, pero no deben necesariamente alimentar el modelo predictivo. También convierte variables categóricas como tipo_tasa_ref_homologado, Calificación, Plazo_Agrupado, Emisor y Bucket_Ancla en categorías, dejando la base ordenada temporalmente y lista para entrenamiento.

En la ejecución reportada, el preprocesamiento generó el archivo dataset_ml_limpio.parquet con 90 columnas finales. La salida muestra además una cobertura amplia de las variables ancladas:

Ancla_Tasa_Fija cuenta con 431.072 valores no nulos; Residual_Tasa_Fija, con 421.774; Spread_Indexado, con 9.334; Ancla_Spread_Indexado, con 9.298; y Residual_Spread, con 9.298. Estos resultados indican que la base final conserva una cobertura alta para tasa fija y una muestra más reducida para instrumentos indexados, lo cual debe tenerse en cuenta en la interpretación posterior de los modelos.

Figura 9

Resultados Pipeline 5_Preprocesamiento_ML.py

```
=====
PREPROCESAMIENTO FINAL PARA MACHINE LEARNING ANCLADO
=====
[~] Filas brutas en Master: 1,673,550
[~] Filas tras deduplicación: 431,191
[~] Columnas eliminadas por leak o ruido: ['Secuencia', 'Tipo', 'Archivo_Origen', 'Dias_Rezago_Precia', 'fecha_corte_precia']
[~] Variables categóricas: ['tipo_tasa_ref_homologado', 'Calificación', 'Plazo_Agrupado', 'Emisor', 'Bucket_Ancla']
=====
[OK] Dataset anclado y guardado en: dataset_ml_limpio.parquet
[!] Dimensiones finales: (431141, 90)
[~] No nulos Spread_Indexado: 9,334
[~] No nulos Ancla_Tasa_Fija: 431,072
[~] No nulos Ancla_Spread_Indexado: 9,298
[~] No nulos Residual_Tasa_Fija: 421,774
[~] No nulos Residual_Spread: 9,298
```

Nota. Elaboración propia a partir de la ejecución del script.

Capa 8. Modelado temporal supervisado y selección de modelos

Una vez construido el dataset limpio, se ejecutó la etapa de modelado mediante el script 6_Modelado_ML_Avanzado.py. Esta capa tuvo como objetivo entrenar, comparar y seleccionar modelos supervisados de regresión para distintos escenarios de pricing: tasa fija, spread IBR, spread IPC y basis general.

El script parte del archivo dataset_ml_limpio.parquet y define cuatro configuraciones de target. El primer escenario, TASA_FIJA, filtra instrumentos de tasa fija y utiliza como variable objetivo la TIR. El segundo, SPREAD_IBR, filtra instrumentos indexados al IBR y utiliza como objetivo Spread_Indexado. El tercero, SPREAD_IPC, filtra instrumentos indexados al IPC y también utiliza Spread_Indexado. El cuarto, BASIS_GENERAL, utiliza toda la base disponible y toma como objetivo Prima_Emision_Basis.

Esta separación por escenarios es relevante porque el pricing de CDTs no es homogéneo entre modalidades. Una emisión a tasa fija se comporta de manera distinta a una emisión indexada al IBR o al IPC; por tanto, entrenar escenarios diferenciados permite ajustar el modelo a la lógica financiera de cada familia de instrumentos.

Antes del entrenamiento, el script excluye variables que podrían contaminar la predicción, como Tasa, TIR, Spread_Indexado, Prima_Emision_Basis, ISIN, Nemo, fechas de emisión y vencimiento, y campos operativos de tipo de tasa. Esta exclusión es fundamental para controlar leakage, porque evita que el modelo use directamente el target o variables demasiado cercanas al resultado que se busca predecir.

El proceso de entrenamiento utiliza una ventana temporal de los últimos 720 días respecto a la fecha máxima de emisión disponible. Luego ordena la información por fecha_emision y aplica un split temporal 80/20 mediante train_test_split con shuffle=False. Este punto corrige una preocupación metodológica importante: la evaluación no mezcla aleatoriamente observaciones pasadas y futuras, sino

que conserva el orden temporal, replicando mejor las condiciones reales de uso. Además, la búsqueda de hiperparámetros utiliza TimeSeriesSplit con cuatro particiones, lo que refuerza la validación temporal del proceso.

El script también incorpora dos controles adicionales. Primero, aplica winsorización al target entre los percentiles 1 % y 99 %, reduciendo el impacto de valores extremos. Segundo, asigna pesos por recencia mediante una función exponencial, de modo que las observaciones más recientes tengan mayor peso en el cálculo de métricas. Esta decisión es financieramente razonable porque las condiciones de mercado recientes suelen ser más relevantes para una herramienta de pricing que observaciones muy antiguas.

En cuanto al procesamiento de variables, el script utiliza un ColumnTransformer. Las variables numéricas se imputan con la mediana y se escalan con StandardScaler. Las variables categóricas se imputan con la categoría “Desconocido” y se codifican mediante OneHotEncoder. Esta estructura permite integrar variables numéricas y categóricas dentro de un mismo pipeline reproducible.

El torneo de modelos incluye cinco algoritmos: Ridge, ElasticNet, Random Forest, XGBoost y LightGBM. Adicionalmente, se entrena un ensamble tipo Stacking con XGBoost, LightGBM y Random Forest como modelos base, y Ridge como estimador final. Para cada modelo se realiza búsqueda de hiperparámetros mediante GridSearchCV, usando como métrica de optimización el RMSE negativo. Al finalizar, el script guarda los modelos entrenados en formato joblib, los resultados en CSV, metadatos en JSON y dashboards de evaluación con comparación de RMSE y gráfico real vs. predicho.

Resultados del modelado temporal

Los resultados muestran diferencias importantes entre escenarios. En tasa fija, el desempeño fue alto y estable. El mejor modelo fue ElasticNet, con RMSE de 0,000845, MAE de 0,000321 y R^2 de 0,9886. Ridge, Random Forest y LightGBM también presentaron resultados competitivos. Esto sugiere

que las tasas fijas tienen una estructura explicable por las variables construidas, especialmente por curvas, anclas temporales, plazo y señales de mercado.

En el escenario SPREAD_IBR, el mejor modelo fue Ridge, con RMSE de 0,000665, MAE de 0,000475 y R^2 de 0,9804. Sin embargo, la muestra fue mucho más pequeña: 1.240 observaciones de entrenamiento y 311 de prueba. Los modelos basados en árboles presentaron R^2 negativo, lo que indica sobreajuste o dificultad para generalizar en una muestra reducida. Por tanto, para IBR el resultado favorece modelos lineales regularizados, más simples y estables.

En el escenario SPREAD_IPC, el mejor modelo fue ElasticNet, con RMSE de 0,001610, MAE de 0,001151 y R^2 de 0,7695. Aunque el desempeño es aceptable, debe interpretarse con cautela porque la muestra es aún más pequeña: 433 observaciones de entrenamiento y 109 de prueba. En este caso, los modelos más complejos no superaron al modelo regularizado, lo que refuerza la necesidad de usar parsimonia cuando la disponibilidad de datos es limitada.

En el escenario BASIS_GENERAL, el mejor modelo fue XGBoost, con RMSE de 0,003363, MAE de 0,002528 y R^2 de 0,8855. A diferencia de los escenarios indexados, aquí la muestra fue amplia: 187.163 observaciones de entrenamiento y 46.791 de prueba. El buen desempeño de XGBoost sugiere que el basis general incorpora relaciones no lineales entre indexador, plazo, curvas, emisores, anclas temporales y señales competitivas.

Las métricas de desempeño presentadas en la Tabla 21 corresponden al entrenamiento y validación ejecutados con corte al 03 de mayo de 2026, usando la ventana temporal, los filtros de calidad, los controles de leakage y la separación cronológica descritos en la capa de modelado.

Tabla 21

Resultados del Modelado

Escenario	Mejor modelo	RMSE	MAE	R ²	Entrenamiento	Prueba
Tasa fija	ElasticNet	0,000845	0,000321	0,9886	185.488	46.373
Spread IBR	Ridge	0,000665	0,000475	0,9804	1.240	311
Spread IPC	ElasticNet	0,001610	0,001151	0,7695	433	109
Basis general	XGBoost	0,003363	0,002528	0,8855	187.163	46.791

Nota. Elaboración propia a partir de la ejecución del script.

Un hallazgo importante es que el modelo Stacking no fue ganador en ninguno de los escenarios. Aunque conceptualmente un ensamble puede mejorar el desempeño, en esta ejecución presentó resultados inferiores frente a modelos individuales. Por tanto, Stacking se conserva como parte del torneo de modelos evaluado, pero no como modelo operativo principal.

Interpretación metodológica de los resultados

Los resultados respaldan una estrategia diferenciada por tipo de target. Para tasa fija, ElasticNet ofrece un equilibrio entre precisión, estabilidad e interpretabilidad. Para IBR e IPC, los modelos lineales regularizados son preferibles debido al tamaño reducido de la muestra. Para el basis general, XGBoost captura mejor las relaciones no lineales presentes en una base amplia y heterogénea.

Desde CRISP-DM, esta capa corresponde a las fases de modelado y evaluación. El proceso no se limita a entrenar un algoritmo, sino que compara alternativas, usa validación temporal, controla fuga de información, pondera por recencia, guarda metadatos y produce artefactos reproducibles. Esto fortalece la trazabilidad del modelo y permite justificar por qué se selecciona un modelo distinto según el escenario de pricing.

Capa 9. Interpretabilidad, diagnóstico de validación y análisis de drivers del modelo

Después del entrenamiento y selección de modelos por escenario, se desarrolló una capa de interpretabilidad mediante el script 7_Interpretabilidad.py. Esta etapa tuvo como propósito explicar el comportamiento de los modelos ganadores, identificar las variables con mayor influencia predictiva,

evaluar la calidad de ajuste fuera de muestra y generar artefactos visuales que facilitaran la lectura técnica y de negocio de los resultados.

El script parte de los modelos previamente entrenados y guardados en formato joblib, junto con los archivos de resultados y metadatos generados en la etapa de modelado. Para cada escenario — TASA_FIJA, SPREAD_IBR, SPREAD_IPC y BASIS_GENERAL— el proceso identifica automáticamente el modelo ganador a partir del menor RMSE registrado en los archivos de resultados. Luego reconstruye el mismo conjunto de validación temporal usado en la fase de modelado, carga el pipeline ganador y calcula nuevamente las predicciones sobre la muestra de prueba.

Esta reconstrucción es metodológicamente importante porque permite que la interpretabilidad se realice sobre el mismo enfoque temporal utilizado en el entrenamiento. Es decir, los gráficos no explican el desempeño sobre toda la muestra histórica, sino sobre el tramo de validación más reciente, que representa mejor el uso real del modelo para decisiones futuras de pricing.

Métodos de interpretabilidad implementados

La primera técnica utilizada es la importancia por permutación. Esta técnica mide cuánto se deteriora el desempeño del modelo cuando una variable es alterada aleatoriamente, manteniendo las demás constantes. En términos prácticos, si al permutar una variable el error del modelo aumenta de forma importante, se interpreta que dicha variable aporta información predictiva relevante. El script calcula esta importancia usando `permutation_importance`, con diez repeticiones y RMSE como criterio de evaluación.

La segunda técnica corresponde a los coeficientes agregados, disponible cuando el modelo ganador es lineal, como Ridge o ElasticNet. Dado que el pipeline transforma variables categóricas mediante One Hot Encoding y escala variables numéricas, el script mapea las variables transformadas hacia su variable original y agrupa los coeficientes por variable base. Esto permite interpretar la

dirección y magnitud relativa del efecto, aunque con la salvedad de que la lectura debe hacerse dentro del espacio transformado del modelo.

La tercera técnica es la importancia nativa del modelo, disponible para modelos basados en árboles o boosting, como XGBoost, LightGBM o Random Forest. Esta importancia refleja el aporte interno de cada variable dentro de la estructura del modelo. En el caso del escenario BASIS_GENERAL, donde el ganador fue XGBoost, esta técnica permitió complementar la importancia por permutación con una segunda lectura basada en la lógica interna del algoritmo.

Adicionalmente, el script genera cuatro tipos de diagnóstico visual para cada escenario: dashboard de predicción, matriz de correlación, matriz de deciles de regresión y archivos CSV con predicciones, métricas e importancia de variables. El dashboard de predicción incluye evolución agregada del valor real y predicho, gráfico real versus predicho, error absoluto por decil y distribución de residuales. La matriz de deciles permite evaluar si el modelo no solo predice niveles cercanos, sino si también ordena correctamente las observaciones de menor a mayor tasa, spread o basis.

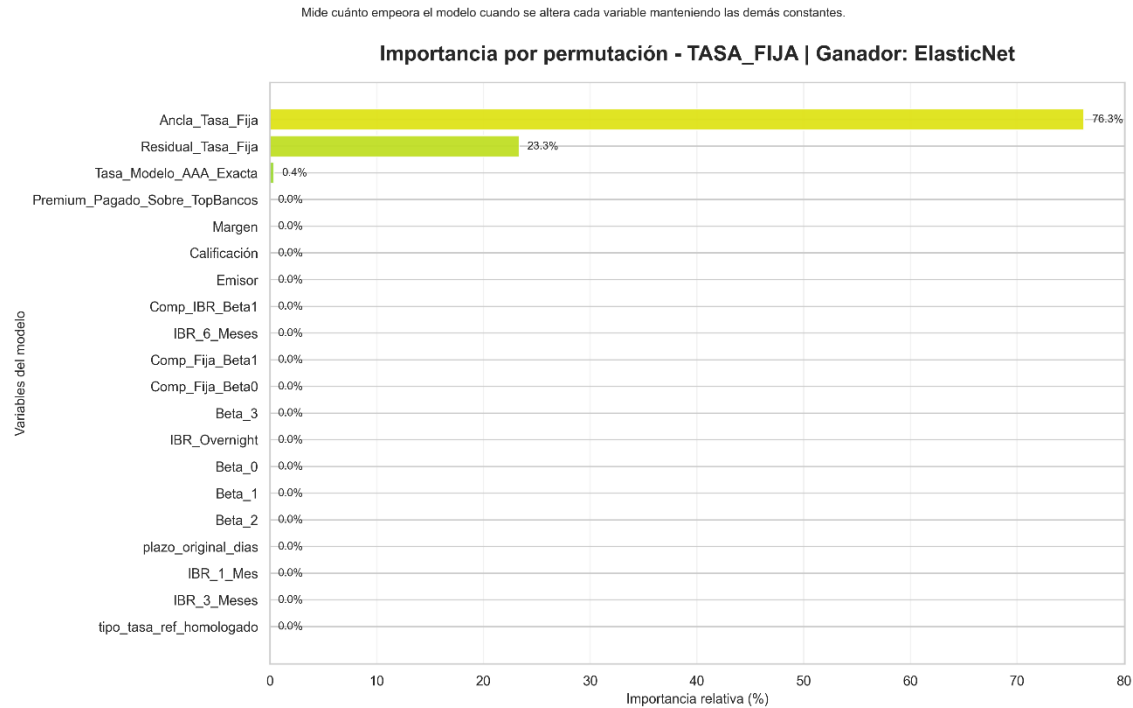
Resultados de interpretabilidad por escenario

Tasa Fija

En el escenario de tasa fija, el modelo ganador fue ElasticNet. Los resultados muestran que el desempeño está explicado principalmente por Ancla_Tasa_Fija y Residual_Tasa_Fija. En la importancia por permutación, Ancla_Tasa_Fija representa aproximadamente el 76,3 % de la importancia relativa, mientras que Residual_Tasa_Fija representa cerca del 23,3 %. Esto indica que el modelo depende principalmente de la referencia histórica reciente por tipo de tasa, plazo y calificación, complementada por el componente residual respecto a dicha ancla.

Figura 10

Importancia por Permutación Para el Escenario de Tasa Fija



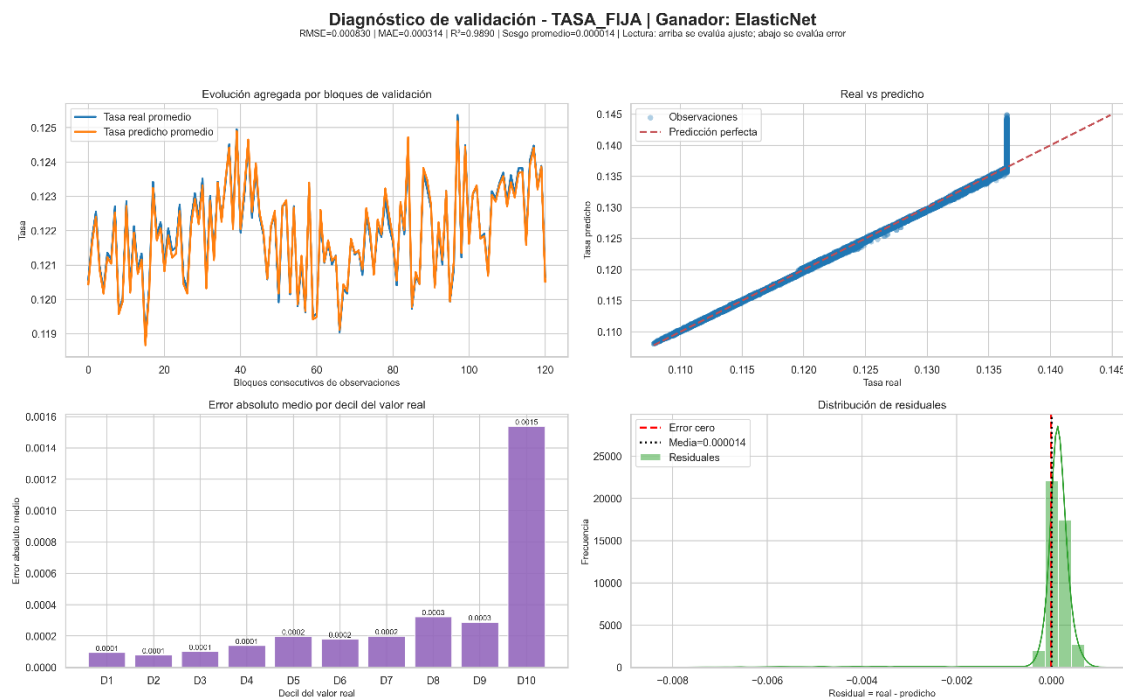
Lectura: una mayor barra indica que alterar esa variable deteriora más el desempeño del modelo. La importancia es relativa y no implica causalidad económica.

Nota. *Elaboración propia a partir de la ejecución del script.*

El diagnóstico visual de tasa fija muestra un ajuste muy fuerte. El gráfico real versus predicho presenta observaciones cercanas a la diagonal, los residuales se concentran alrededor de cero y la matriz de deciles muestra un acierto en diagonal cercano al 96,1 %. Esto significa que el modelo no solo estima adecuadamente el nivel de tasa, sino que también ordena correctamente la mayoría de las observaciones por rango relativo. Este escenario se presenta como el caso de mejor desempeño y mayor estabilidad dentro del torneo de modelos.

Figura 11

Diagnóstico de Validación Para el Escenario de Tasa Fija



Nota metodológica: un buen modelo presenta puntos cercanos a la diagonal en 'Real vs predicho', errores bajos por decil y residuales concentrados alrededor de cero sin sesgo sistemático.

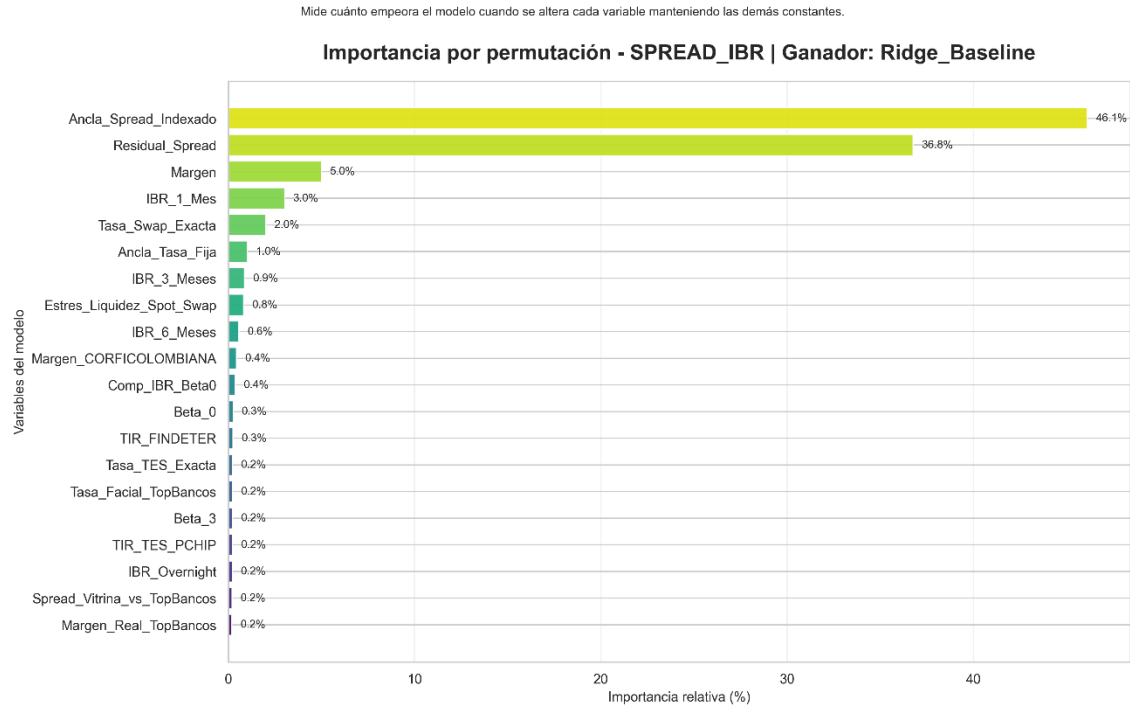
Nota. Elaboración propia a partir de la ejecución del script.

IBR

En el escenario SPREAD_IBR, el modelo ganador fue Ridge. La importancia por permutación muestra que Ancla_Spread_Indexado y Residual_Spread son las variables más relevantes, con participaciones aproximadas de 46,1 % y 36,8 %, respectivamente. También aparecen variables como Margen, IBR_1_Mes, Tasa_Swap_Exacta, IBR_3_Meses, Estres_Liquidez_Spot_Swap e indicadores de curvas. Esta composición es financieramente coherente, porque el spread IBR depende de una

referencia histórica de mercado, del comportamiento reciente del spread y de señales asociadas al mercado monetario y a la curva swap.

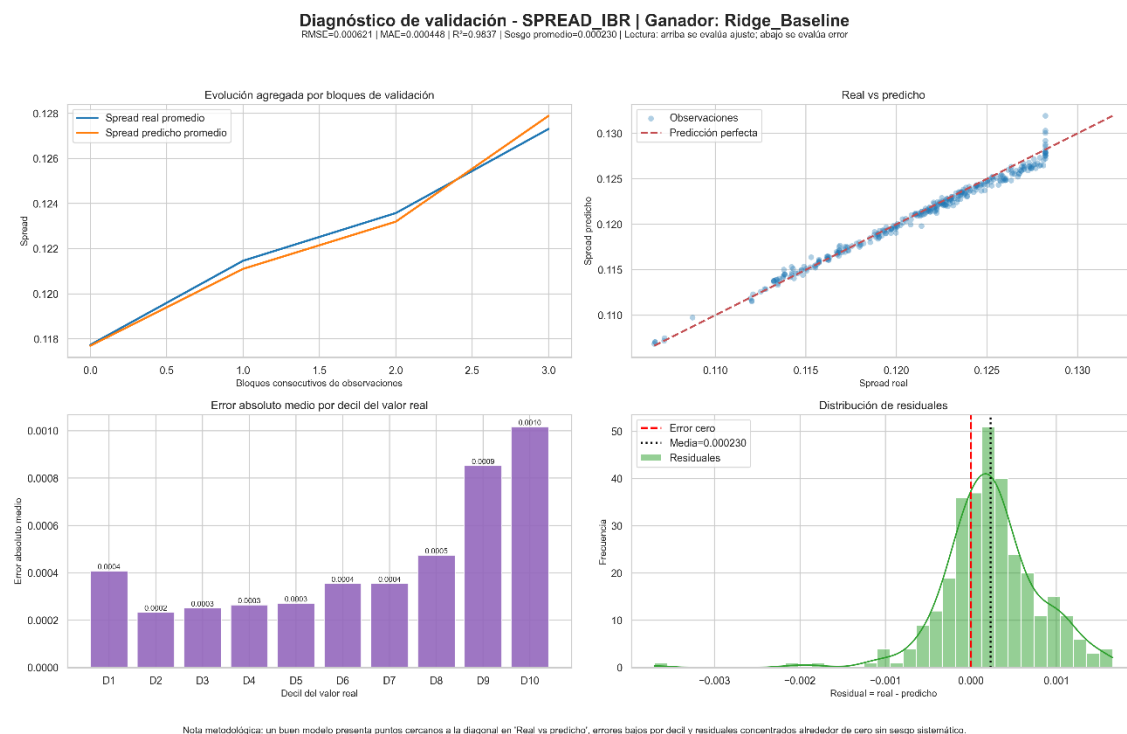
Figura 12
Importancia por Permutación Para el Escenario de IBR



Lectura: una mayor barra indica que alterar esa variable deteriora más el desempeño del modelo. La importancia es relativa y no implica causalidad económica.

Nota. Elaboración propia a partir de la ejecución del script.

El diagnóstico de validación de IBR también es sólido, aunque debe interpretarse con cautela por el menor tamaño de muestra frente a tasa fija. La matriz de deciles muestra un acierto en diagonal cercano al 80,7 %, lo cual indica una buena capacidad de ordenamiento relativo. Además, el gráfico real versus predicho conserva una relación cercana a la diagonal, con residuales relativamente centrados.

Figura 13**Diagnóstico de Validación Para el Escenario de IBR**

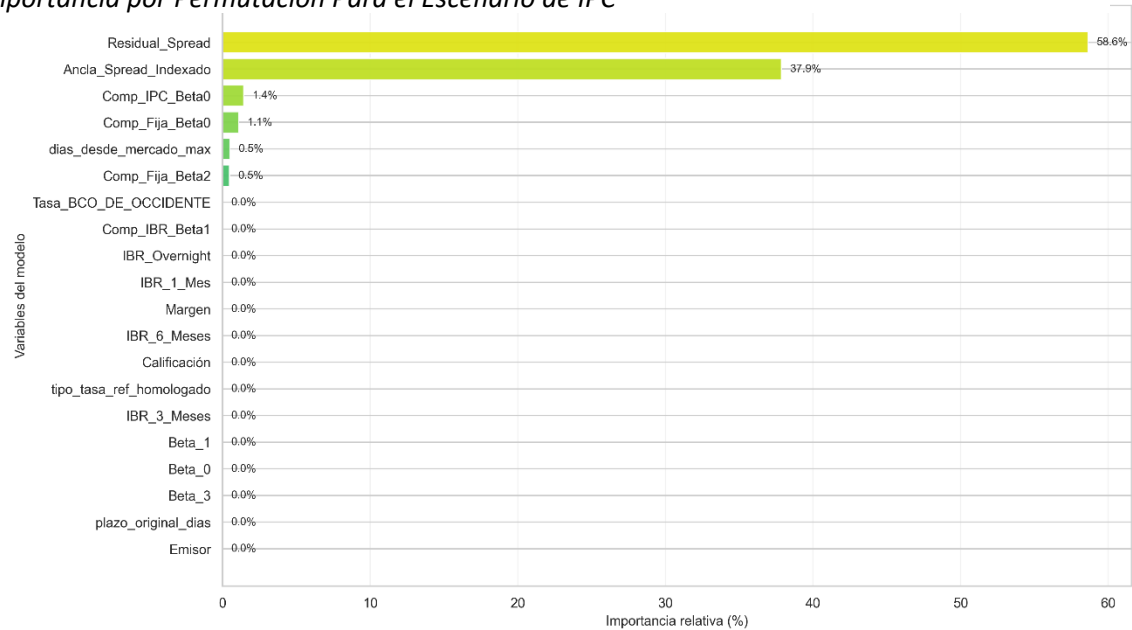
Nota. Elaboración propia a partir de la ejecución del script.

IPC

En el escenario SPREAD_IPC, el modelo ganador fue ElasticNet. La importancia por permutación se concentra principalmente en Residual_Spread y Ancla_Spread_Indexado, con participaciones aproximadas de 58,6 % y 37,9 %. Esto indica que el modelo se apoya de forma dominante en variables ancladas y residuales, más que en señales amplias de mercado. El desempeño es aceptable, pero menos robusto que en tasa fija e IBR. La matriz de deciles muestra un acierto en diagonal cercano al 45 %, lo que evidencia mayor dispersión en la capacidad de ordenamiento.

Figura 14

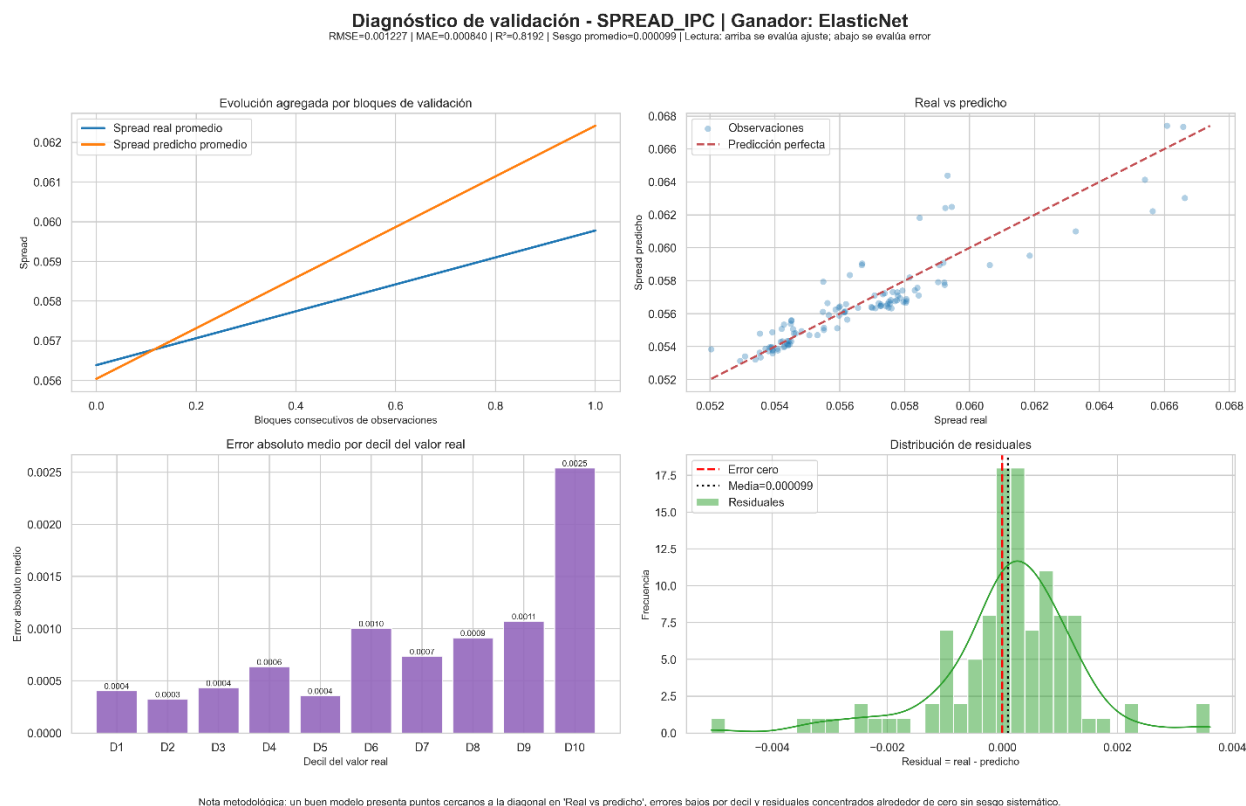
Importancia por Permutación Para el Escenario de IPC



Lectura: una mayor barra indica que alterar esa variable deteriora más el desempeño del modelo. La importancia es relativa y no implica causalidad económica.

Nota. Elaboración propia a partir de la ejecución del script.

Este resultado no invalida el escenario IPC, pero requiere una interpretación prudente. La muestra de instrumentos IPC es reducida y el modelo parece capturar parcialmente la estructura del spread, pero con menor estabilidad. El modelo IPC se interpreta como una aproximación analítica funcional, aunque requiere mayor profundidad histórica para robustecer su uso operativo.

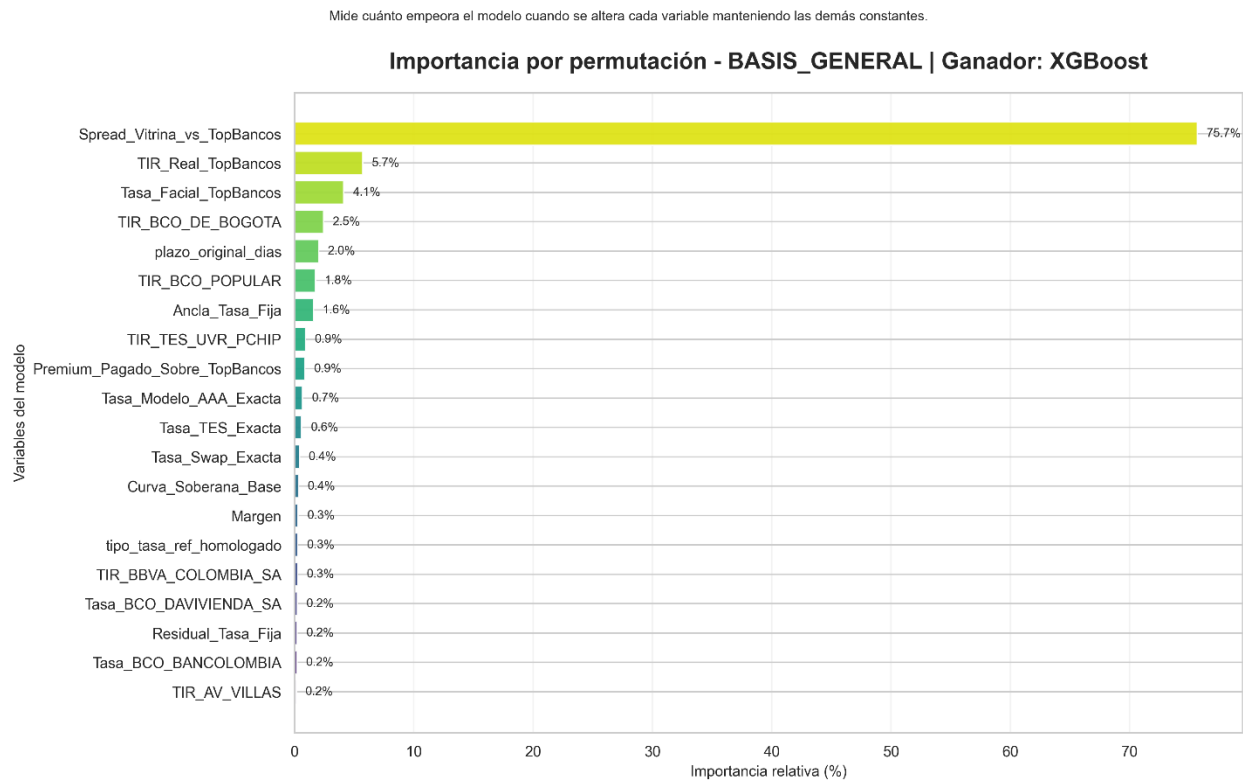
Figura 15**Diagnóstico de Validación Para el Escenario de IPC**

Nota. Elaboración propia a partir de la ejecución del script.

Basis

En el escenario BASIS_GENERAL, el modelo ganador fue XGBoost. La importancia por permutación muestra una concentración muy fuerte en Spread_Vitrina_vs_TopBancos, con una participación aproximada de 75,7 %. También aparecen variables competitivas como TIR_Real_TopBancos, Tasa_Facial_TopBancos, TIR_BCO_DE_BOGOTA, TIR_BCO_POPULAR, además de plazo_original_dias, Ancla_Tasa_Fija, TIR_TES_UVR_PCHIP, Premium_Pagado_Sobre_TopBancos y Tasa_Modelo_AAA_Exacta.

Figura 16
Importancia por Permutación Para el Escenario de Basis



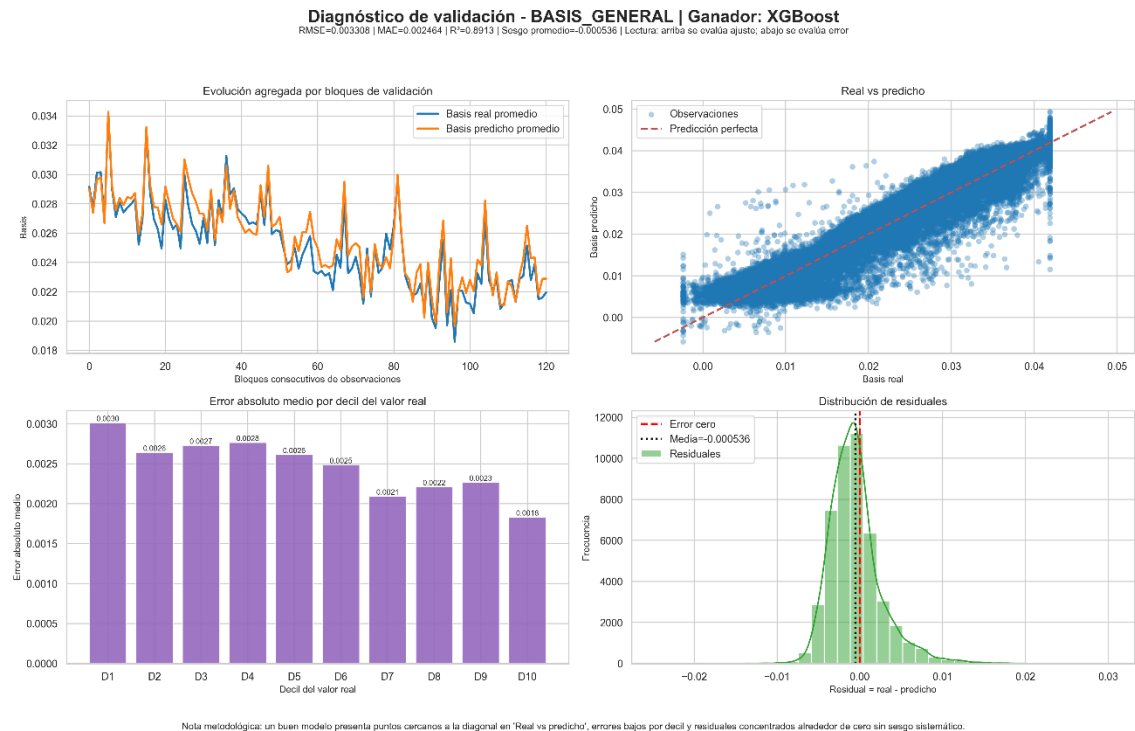
Lectura: una mayor barra indica que alterar esa variable deteriora más el desempeño del modelo. La importancia es relativa y no implica causalidad económica.

Nota. Elaboración propia a partir de la ejecución del script

Este hallazgo es relevante para la interpretación financiera del modelo, porque muestra que el basis general está fuertemente influenciado por variables de posicionamiento competitivo frente al grupo de bancos comparables. En otras palabras, el modelo no solo usa curvas soberanas o variables macro, sino que aprende de la relación entre la tasa de vitrina, la tasa promedio de competidores y las condiciones efectivas observadas en el mercado.

Figura 17

Diagnóstico de Validación Para el Escenario de Basis



Nota. Elaboración propia a partir de la ejecución del script.

La importancia nativa de XGBoost muestra una lectura complementaria:

dias_desde_mercado_max aparece como una variable relevante dentro de la estructura interna del modelo. Esto sugiere que la recencia de la información contribuye al aprendizaje del algoritmo. Esta importancia nativa no se interpreta como causalidad económica directa, sino como aporte interno del modelo.

Tabla 22

Resumen Interpretabilidad

Escenario	Modelo ganador	Técnicas de interpretabilidad relevantes	Lectura principal
Tasa fija	ElasticNet	Importancia por permutación, coeficientes agregados, residuales y deciles	El modelo depende principalmente de Ancla_Tasa_Fija y Residual_Tasa_Fija; presenta alto ajuste y muy buen ordenamiento.

Spread IBR	Ridge	Importancia por permutación, coeficientes agregados, correlaciones y deciles	El spread se explica por ancla indexada, residual, margen, IBR y curva swap; desempeño sólido con muestra reducida.
Spread IPC	ElasticNet	Importancia por permutación, coeficientes agregados y diagnóstico visual	El modelo depende de ancla y residual, pero presenta menor capacidad de ordenamiento; debe usarse con cautela.
Basis general	XGBoost	Importancia por permutación, importancia nativa, residuales y deciles	El basis general se explica principalmente por variables competitivas frente al top de bancos y por señales de recencia.

Nota. Elaboración propia a partir de la ejecución del script

Capa 10. Despliegue funcional mediante aplicación Streamlit

La última capa del pipeline corresponde al prototipo funcional de despliegue, desarrollado en el script 8_App_Streamlit.py. Esta aplicación permite integrar en una interfaz de usuario los resultados del dataset limpio, el torneo de modelos, los artefactos de interpretabilidad y un simulador de pricing de emisión. Desde la perspectiva de CRISP-DM, esta capa se relaciona con la fase de despliegue, dado que traduce los modelos entrenados en una herramienta consultable por usuarios de negocio.

La aplicación está construida en Streamlit y se organiza en cuatro pestañas principales: Datos, Torneo, Interpretabilidad y Motor de Pricing. Esta estructura permite que el usuario no solo consulte una recomendación final, sino que también pueda revisar la base de datos utilizada, los modelos comparados, los resultados de interpretabilidad y la lógica de simulación por escenario.

Pestaña Datos

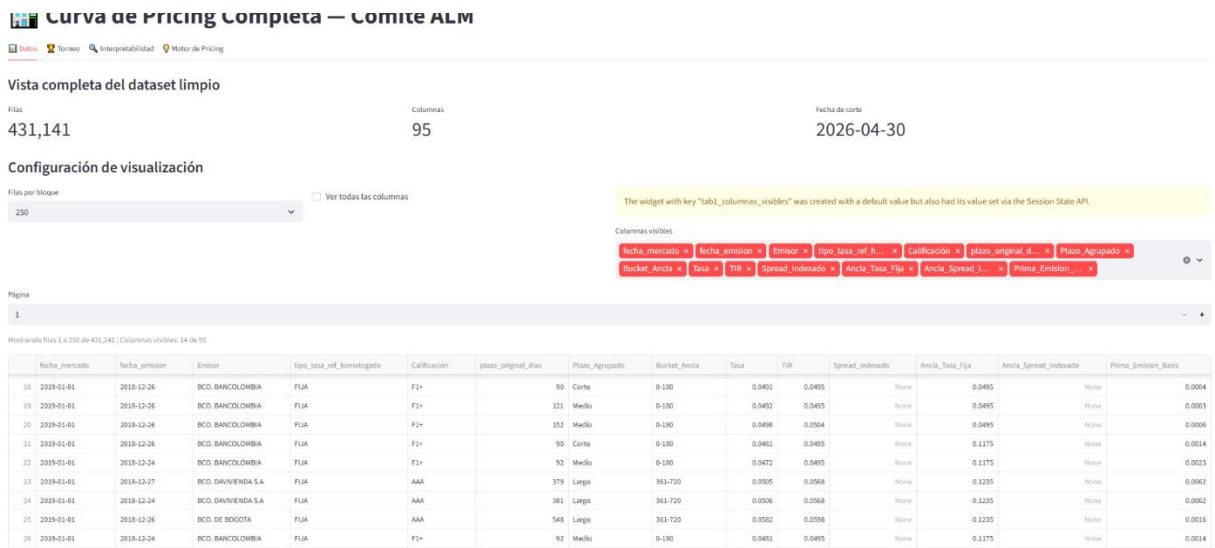
La pestaña Datos carga el archivo dataset_ml_limpio.parquet y presenta una vista navegable del dataset final. La aplicación muestra métricas generales como número de filas, número de columnas y fecha de corte. También permite seleccionar columnas visibles, paginar registros y revisar variables clave como fecha de mercado, emisor, tipo de tasa, calificación, plazo, tasa, TIR, spread, anclas y basis.

Esta pestaña cumple una función de trazabilidad y validación operativa, porque permite inspeccionar la base que alimenta los modelos sin tener que abrir directamente archivos Parquet o

ejecutar código en Python. Además, incluye una visualización de distribución por tipo de tasa, útil para identificar la concentración de la muestra entre tasa fija, IBR e IPC.

Figura 18

Aplicación Pestaña Datos



Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Pestaña Torneo

La pestaña Torneo lee los archivos resultados_{escenario}.csv y metadata_{escenario}.json generados en la etapa de modelado. Para cada escenario, presenta el ranking de modelos, un gráfico de RMSE, el modelo ganador, la métrica de error, la ventana temporal usada y los tamaños de entrenamiento y prueba.

Esta funcionalidad es importante porque convierte el proceso de selección de modelos en una salida auditable para el usuario. En lugar de mostrar únicamente el modelo final, la app permite observar qué alternativas fueron evaluadas y por qué se seleccionó un modelo ganador. Esto fortalece la transparencia del motor y evita que la recomendación se perciba como una “caja negra”.

Figura 19**Aplicación Pestaña Torneo – Escenario Tasa Fija****Curva de Pricing Completa – Comité ALM**

Datos Torneo Interpretabilidad Motor de Pricing

Ranking de modelos

Escenario

TASA_FIJA

Modelo	RMSE	MSE	R2	Target	Mejores_Params	N_Combinaciones
0 ElasticNet		0.0008	0.0003	0.9886 TIR	["modelo__alpha": 0.0005, "modelo__l1_ratio": 0.1]	9
1 Ridge_Baseline		0.0009	0.0002	0.988 TIR	["modelo__alpha": 0.1]	4
2 RandomForest		0.001	0.0007	0.9852 TIR	["modelo__max_depth": null, "modelo__min_samples_leaf": 2, "modelo__n_estimators": 200]	18
3 LightGBM		0.0013	0.0009	0.9742 TIR	["modelo__learning_rate": 0.05, "modelo__min_child_samples": 10, "modelo__n_estimators": 250, "modelo__n_threads": 4]	54
4 XGBoost		0.0013	0.0011	0.9723 TIR	["modelo__colsample_bytree": 0.8, "modelo__learning_rate": 0.05, "modelo__max_depth": 4, "modelo__n_estimators": 200]	72
5 Stacking (XGB+LGBM+RF)		0.0008	0.0006	0.7722 TIR	["estimator": ["XGBBoost", "LightGBM", "RandomForest"], "final_estimator": "Ridge(alpha=1.0)", "cv": 5]	1

RMSE por Modelo - TASA_FIJA*Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.***Figura 20****Aplicación Pestaña Torneo – Escenario IBR****Curva de Pricing Completa – Comité ALM**

Datos Torneo Interpretabilidad Motor de Pricing

Ranking de modelos

Escenario

SPREAD_IBR

Modelo	RMSE	MSE	R2	Target	Mejores_Params	N_Combinaciones
0 Ridge_Baseline		0.0007	0.0005	0.9804 Spread_Indeado	["modelo__alpha": 0.1]	4
1 ElasticNet		0.0008	0.0006	0.9742 Spread_Indeado	["modelo__alpha": 0.0005, "modelo__l1_ratio": 0.1]	9
2 LightGBM		0.005	0.007	-1.7992 Spread_Indeado	["modelo__learning_rate": 0.03, "modelo__min_child_samples": 10, "modelo__n_estimators": 250, "modelo__n_threads": 4]	54
3 RandomForest		0.0084	0.0075	-2.1258 Spread_Indeado	["modelo__max_depth": null, "modelo__min_samples_leaf": 2, "modelo__n_estimators": 200]	18
4 XGBoost		0.0098	0.009	-3.2433 Spread_Indeado	["modelo__colsample_bytree": 0.8, "modelo__learning_rate": 0.05, "modelo__max_depth": 5, "modelo__n_estimators": 200]	72
5 Stacking (XGB+LGBM+RF)		0.0219	0.0214	-20.2504 Spread_Indeado	["estimator": ["XGBBoost", "LightGBM", "RandomForest"], "final_estimator": "Ridge(alpha=1.0)", "cv": 5]	1

RMSE por Modelo - SPREAD_IBR*Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.*

Figura 21**Aplicación Pestaña Torneo – Escenario IPC****Curva de Pricing Completa – Comité ALM**

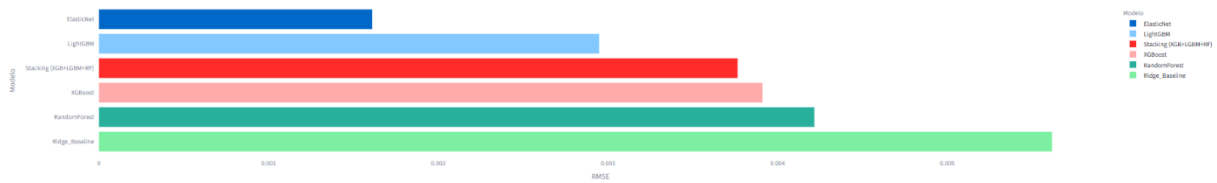
Datos
 Torneo
 Interpretabilidad
 Motor de Pricing

Ranking de modelos

Escenario

SPREAD_IPC

	Modelo	RMSE	MAE	R2	Target	Mejores_Params	N_Combinaciones
0	ElasticNet	0.0016	0.0012	0.7095	Spread_Indexado	["modelo__alpha": 0.0005, "modelo__l1_ratio": 0.1]	9
1	LightGBM	0.0029	0.0021	0.2266	Spread_Indexado	["modelo__learning_rate": 0.06, "modelo__min_child_samples": 10, "modelo__n_estimators": 250, "modelo__n_threads": 4, "modelo__num_leaves": 31, "modelo__scale_pos_weight": 1.0, "modelo__subsample": 0.9, "modelo__verbose": -1]	54
2	Stacking (GB+LGBM+RF)	0.0038	0.0031	-4.2681	Spread_Indexado	["estimator": ["XGBoost", "LightGBM", "RandomForest"], "final_estimator": "Ridge(alpha=1.0)", "cv": 3]	1
3	XGBoost	0.0039	0.0026	-6.3603	Spread_Indexado	["modelo__colsample_bytree": 0.6, "modelo__learning_rate": 0.03, "modelo__max_depth": 3, "modelo__n_estimators": 200, "modelo__n_threads": 4, "modelo__num_leaves": 31, "modelo__scale_pos_weight": 1.0, "modelo__subsample": 0.9, "modelo__verbose": -1]	72
4	RandomForest	0.0042	0.0034	-6.5813	Spread_Indexado	["modelo__max_depth": 10, "modelo__min_samples_leaf": 10, "modelo__n_estimators": 200]	18
5	Ridge_Baseline	0.0056	0.0037	-1.8568	Spread_Indexado	["modelo__alpha": 1.0]	4

RMSE por Modelo - SPREAD_IPC

Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Figura 22**Aplicación Pestaña Torneo – Escenario Basis****Curva de Pricing Completa – Comité ALM**

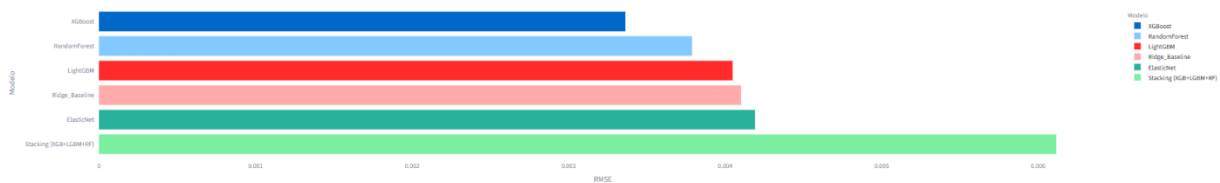
Datos
 Torneo
 Interpretabilidad
 Motor de Pricing

Ranking de modelos

Escenario

BASIS_GENERAL

	Modelo	RMSE	MAE	R2	Target	Mejores_Params	N_Combinaciones
0	XGBoost	0.0034	0.0025	0.8855	Prima_Emission_Basis	["modelo__colsample_bytree": 1.0, "modelo__learning_rate": 0.08, "modelo__max_depth": 5, "modelo__n_estimators": 250, "modelo__n_threads": 4, "modelo__num_leaves": 31, "modelo__scale_pos_weight": 1.0, "modelo__subsample": 0.9, "modelo__verbose": -1]	72
1	RandomForest	0.0038	0.0029	0.9547	Prima_Emission_Basis	["modelo__max_depth": null, "modelo__min_samples_leaf": 2, "modelo__n_estimators": 200]	18
2	LightGBM	0.004	0.0032	0.8341	Prima_Emission_Basis	["modelo__learning_rate": 0.08, "modelo__min_child_samples": 40, "modelo__n_estimators": 250, "modelo__n_threads": 4, "modelo__num_leaves": 31, "modelo__scale_pos_weight": 1.0, "modelo__subsample": 0.9, "modelo__verbose": -1]	54
3	Ridge_Baseline	0.0041	0.0027	0.8297	Prima_Emission_Basis	["modelo__alpha": 1.0]	4
4	ElasticNet	0.0042	0.0029	0.9221	Prima_Emission_Basis	["modelo__alpha": 0.0005, "modelo__l1_ratio": 0.1]	9
5	Stacking (GB+LGBM+RF)	0.0061	0.0053	0.6214	Prima_Emission_Basis	["estimator": ["XGBoost", "LightGBM", "RandomForest"], "final_estimator": "Ridge(alpha=1.0)", "cv": 3]	1

RMSE por Modelo - BASIS_GENERAL

Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Pestaña Interpretabilidad

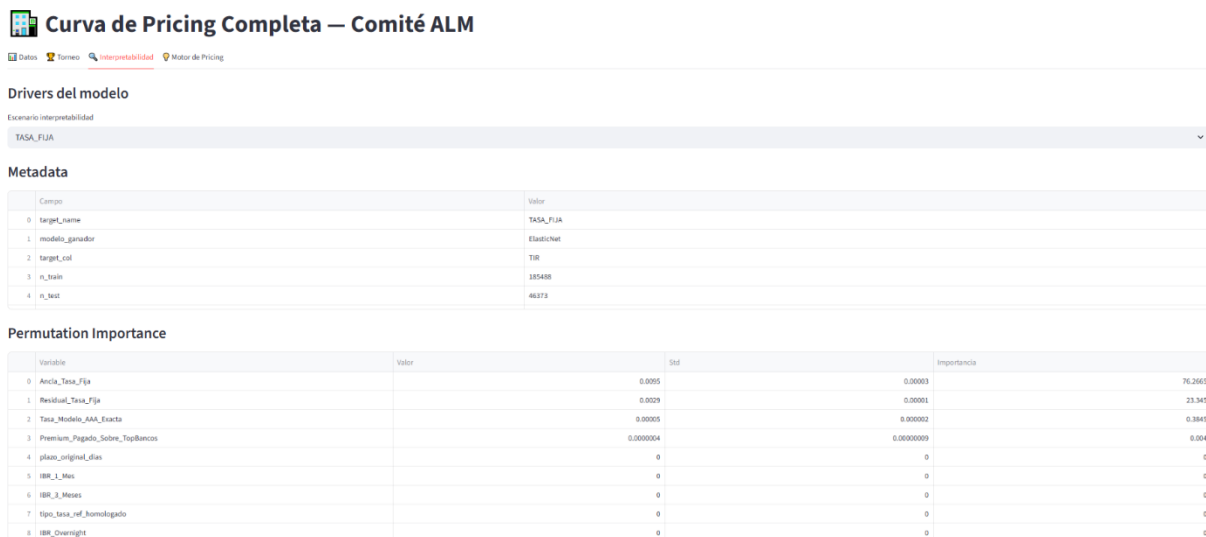
La pestaña Interpretabilidad lee los artefactos generados por el script 7_Interpretabilidad.py.

Para cada escenario, muestra metadata, tablas de importancia por permutación, coeficientes agregados cuando existen, detalle de coeficientes transformados y gráficos como importancia de variables, dashboard de validación, matriz de correlación, matriz de deciles, coeficientes e importancia nativa.

Esta pestaña es clave para el uso gerencial del modelo. Permite que un usuario de tesorería o un comité ALM revise cuáles fueron los principales drivers de una predicción, cómo se comportó el modelo en validación y qué tan estable fue el ajuste. De esta manera, la aplicación no solo entrega una tasa sugerida, sino también evidencia para sustentar la recomendación.

Figura 23

Aplicación Pestaña Interpretabilidad – Escenario Tasa Fija

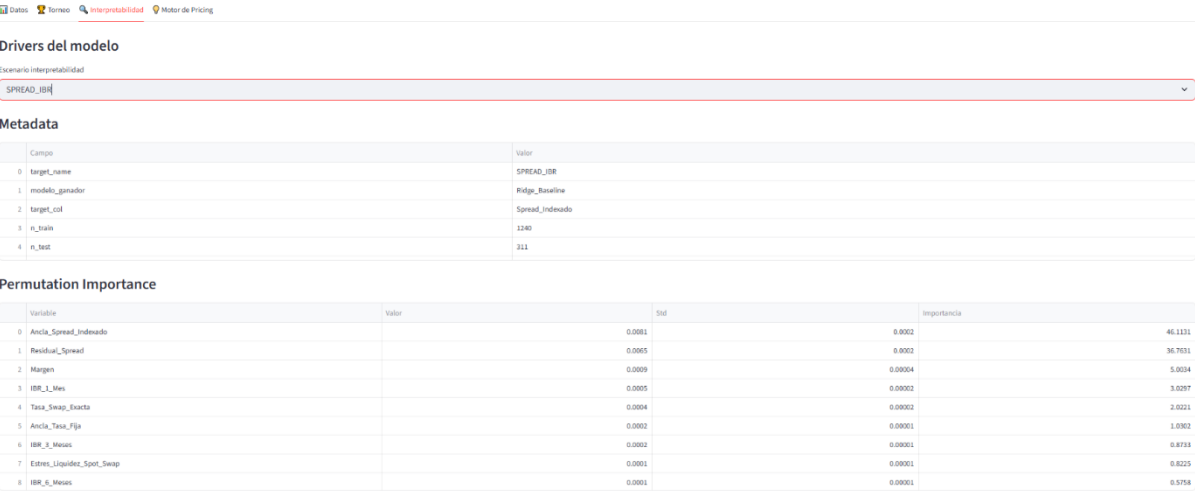


Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Figura 25

Aplicación Pestaña Interpretabilidad – Escenario IBR

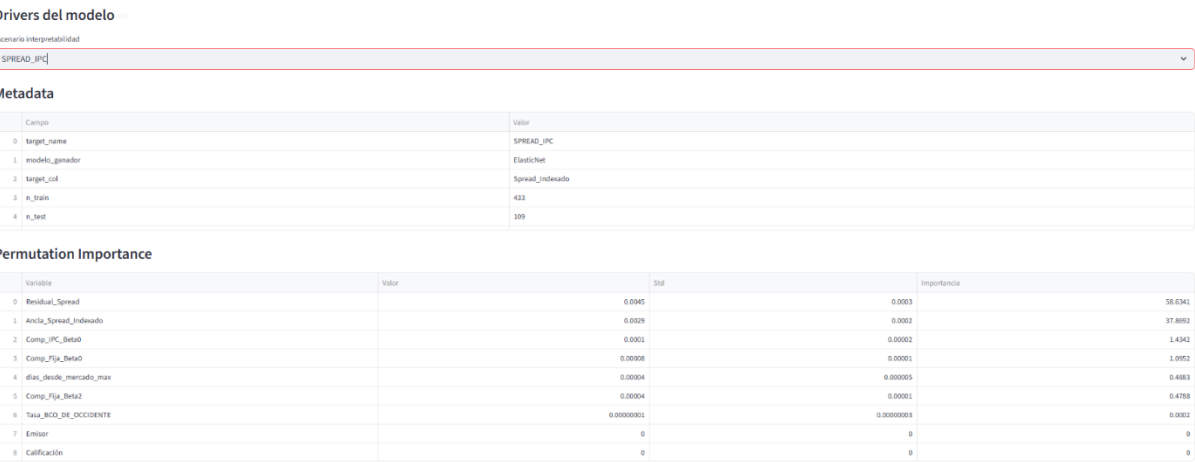
Curva de Pricing Completa – Comité ALM



Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

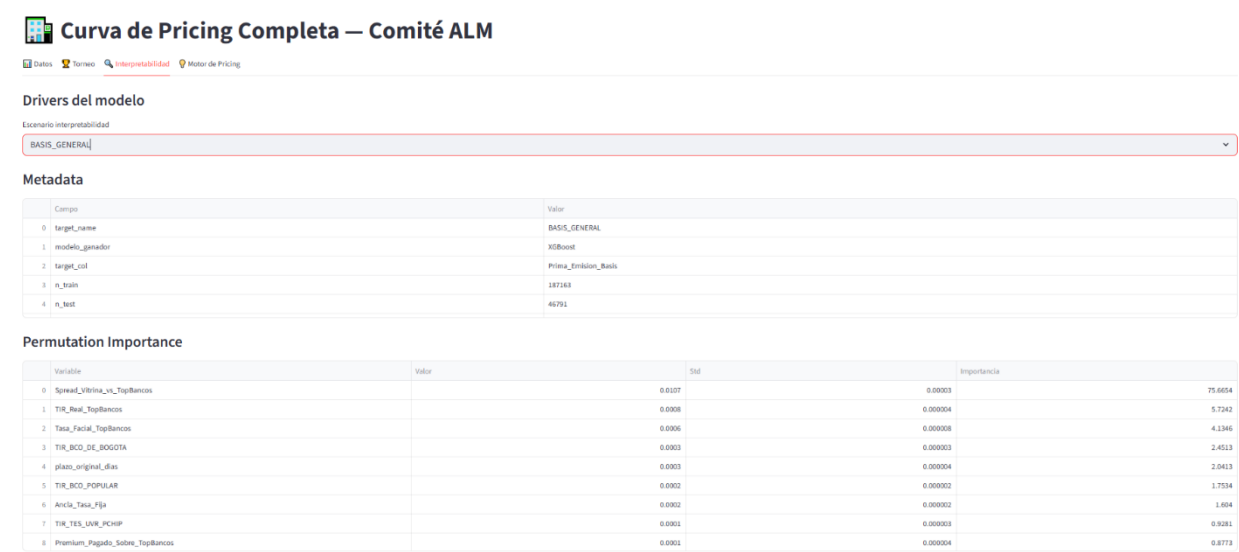
Figura 24

Aplicación Pestaña Interpretabilidad – Escenario IPC



Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Figura 26
Aplicación Pestaña Interpretabilidad – Escenario Basis



Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Pestaña Motor de Pricing

La pestaña Motor de Pricing funciona como un simulador de emisión. El usuario puede seleccionar tipo de tasa —FIJA, IBR o IPC—, plazo en días, emisor, calificación y un choque de mercado en puntos básicos. Con estos insumos, la app construye un registro sintético de entrada a partir de observaciones recientes del dataset limpio, carga el modelo ganador correspondiente y genera una predicción.

La función construir_input selecciona un pool reciente de observaciones comparables por tipo de tasa y plazo. Luego utiliza medianas recientes por bucket para completar variables de mercado, define categorías como Plazo_Agrupado y Bucket_Ancla, asigna emisor y calificación, y aplica choques de mercado sobre variables asociadas a curvas, TES, PCHIP o referencias de mercado. Posteriormente, el modelo ganador genera una recomendación de tasa o spread.

La app también puede construir una curva completa de pricing para plazos estándar: 90, 120, 180, 240, 360, 720, 1080, 1440, 1800 y 2160 días. Para cada plazo, calcula una recomendación para tasa fija, IBR e IPC y la presenta en tabla y gráfico. Esta funcionalidad es útil para transformar el modelo en una herramienta de análisis por estructura de vencimientos, no únicamente en una predicción puntual.

Figura 27

Aplicación Pestaña Motor de Pricing – Cálculo de Precio Según Parámetros

Curva de Pricing Completa – Comité ALM

 Datos  Torneo  Interpretabilidad  Motor de Pricing

Simulador de Pricing de Emisión

Tipo de tasa: FIJA Calificación: AAA

Plazo (días): 360 Choque Tasa de Mercado (bps): 0

Emitor: CORFICOLOMBIANA

 Calcular Precio Técnico  Construir Curva Completa

Modelo ganador: ElasticNet Escenario: TASA_FIJA Bucket: 181-360 RMSE modelo: 0.000845

TIR fija sugerida: 12.68% E.A.

Concepto	Valor
0 Predicción modelo (nivel)	12.68%
1 Ancla usada (referencia)	12.12%
2 TIR final	12.68%

Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Figura 28

Aplicación Pestaña Motor de Pricing – Cálculo Curva Completa

Curva de Pricing Completa – Comité ALM

 Datos  Torneo  Interpretabilidad  Motor de Pricing

Simulador de Pricing de Emisión

Tipo de tasa: FIJA Calificación: AAA

Plazo (días): 360 Choque Tasa de Mercado (bps): 0

Emitor: CORFICOLOMBIANA

 Calcular Precio Técnico  Construir Curva Completa

Curva de Pricing Completa

Plazo (días)	Bucket	Años Aprox.	TIR Fija (% E.A.)	Indicador IBR	Spread IBR	TIR Total (IBR (% Nom))	Indicador IPC	Spread IPC	TIR Total IPC (% E.A.)
90	Corto		0.2 11.31%	11.90%	0.75%	12.70%	5.30%	6.80%	12.15%
1	120 Medio		0.3 11.31%	11.90%	0.75%	12.70%	5.30%	6.80%	12.15%
2	180 Medio		0.5 11.31%	11.90%	0.75%	12.70%	5.30%	6.80%	12.15%
3	240 Medio		0.7 12.68%	12.09%	1.00%	13.09%	4.20%	6.99%	11.19%
4	360 Medio		1 12.68%	12.08%	1.00%	13.08%	4.20%	6.99%	11.19%
5	720 Largo		2 12.83%	11.96%	1.35%	13.31%	5.30%	6.98%	12.28%
6	1080 Muy Largo		3 13.84%	12.55%	1.30%	14.85%	5.73%	6.80%	12.52%
7	1440 Muy Largo		3.9 13.61%	12.67%	1.65%	14.32%	6.05%	6.83%	12.88%
8	1800 Muy Largo		4.9 13.61%	12.63%	1.65%	14.28%	6.05%	6.83%	12.88%
9	2160 Muy Largo		5.9 13.61%	12.60%	1.65%	14.25%	6.05%	6.83%	12.88%

Nota. Elaboración propia a partir de la ejecución del prototipo Streamlit.

Escalabilidad, arquitectura objetivo y operación MLOps

Aunque el prototipo desarrollado en Streamlit se ejecutó inicialmente en un entorno local y controlado, su adopción institucional requiere una arquitectura escalable que permita operación segura, actualización periódica, trazabilidad y monitoreo continuo. Por tanto, la versión productiva del motor no debe depender de la ejecución manual en un equipo local, sino de una arquitectura desplegada en nube privada, nube pública o infraestructura corporativa administrada, con controles de acceso, versionamiento de artefactos, programación automática del pipeline y registro de uso.

Desde una perspectiva MLOps, la solución debe evolucionar hacia un flujo automatizado que integre ingesta de datos, validación de calidad, generación del dataset maestro, entrenamiento o recalibración de modelos, registro de modelos, despliegue controlado, monitoreo de desempeño y alertas de deriva (Google Cloud, 2024). Esta lógica es consistente con las buenas prácticas de MLOps, que promueven la automatización de integración continua, entrega continua y entrenamiento continuo para sistemas de aprendizaje automático (Google Cloud, 2024).

En una arquitectura objetivo, los datos crudos de Precia y Banco de la República deberían almacenarse en una zona de datos crudos; los archivos procesados en Parquet deberían mantenerse en una zona curada; los modelos entrenados deberían registrarse en un repositorio de modelos; y la aplicación Streamlit debería consumir únicamente artefactos validados. Para uso empresarial, Streamlit puede desplegarse en alternativas administradas como Streamlit Community Cloud para pruebas o Streamlit in Snowflake para escenarios corporativos con control de acceso basado en roles (Snowflake Inc, s/f), aunque la decisión final debe alinearse con la arquitectura tecnológica y las políticas de seguridad de Corficolombiana.

Esta evolución permitiría pasar de un prototipo analítico funcional a una solución gobernada. La aplicación conservaría su utilidad para el usuario de mesa, pero el procesamiento, el almacenamiento, el

monitoreo y la actualización quedarían soportados por una infraestructura robusta. En consecuencia, el motor prescriptivo podría operar como una herramienta institucional de apoyo a decisiones de pricing, con trazabilidad de datos, modelos, versiones, usuarios, predicciones y overrides.

Tabla 23

Arquitectura Objetivo Para Escalabilidad y Operación MLOps

Componente	Función en producción	Recomendación
Repositorio de código	Versionar scripts, app y configuración	Git corporativo con revisión de cambios
Orquestación	Ejecutar ingesta, transformación y recalibración	Scheduler corporativo, Airflow, Azure ML Pipelines, Vertex AI Pipelines o equivalente
Almacenamiento	Separar datos crudos, procesados y datasets de modelado	Data lake corporativo o nube aprobada
Registro de modelos	Mantener versiones, métricas y artefactos	MLflow, Azure ML Registry, Vertex AI Model Registry o equivalente
Despliegue	Publicar la app para usuarios autorizados	Servidor corporativo, contenedor Docker o servicio administrado
Seguridad	Controlar acceso y proteger insumos	RBAC, autenticación corporativa y cifrado
Monitoreo	Medir calidad, drift, uso y desempeño	Alertas, logs y métricas periódicas
Gobierno	Aprobar cambios y recalibraciones	Comité Tesorería–Riesgo–Tecnología

Nota. Elaboración propia.

Validación funcional del prototipo con usuarios de mesa

Una vez construido el prototipo funcional en Streamlit, se realizó una validación exploratoria con usuarios vinculados al proceso de pricing, fondeo y gestión ALM. La validación no tuvo como propósito medir impacto financiero definitivo, sino evaluar la utilidad práctica de la herramienta, la claridad de sus salidas, la coherencia de las recomendaciones frente al criterio experto y las condiciones mínimas requeridas para una adopción controlada en la mesa.

La validación funcional posterior se realizó con tres usuarios de ese universo funcional, con propósito exploratorio y no estadístico.

La prueba consistió en una navegación guiada por las cuatro pestañas de la aplicación: Datos, Torneo, Interpretabilidad y Motor de Pricing. Los participantes revisaron escenarios de tasa fija, IBR e IPC en distintos plazos, compararon las recomendaciones del motor frente a su lectura de mercado y evaluaron la utilidad de los drivers presentados por el modelo. Las respuestas fueron analizadas de manera agregada, manteniendo la anonimización de los participantes y el uso estrictamente académico-empresarial de la información.

Tabla 24

Instrumento de Validación Funcional del Prototipo Con Usuarios de la Mesa

Dimensión evaluada	Pregunta aplicada	Propósito
Coherencia de la recomendación	¿La tasa o spread sugerido por la herramienta resulta razonable frente a las condiciones actuales de mercado?	Evaluar alineación con criterio experto.
Utilidad de benchmarks	¿La comparación frente a curvas, top bancos y referencias de mercado aporta valor para decidir una cotización?	Validar utilidad del componente comparativo.
Interpretabilidad	¿Los drivers del modelo ayudan a entender por qué se recomienda una tasa o spread?	Evaluar transparencia de la recomendación.
Trazabilidad	¿La herramienta facilita justificar posteriormente una decisión de pricing?	Evaluar capacidad de auditoría y reconstrucción.
Uso operativo	¿En qué momentos del flujo real de mesa usaría la herramienta?	Identificar puntos de inserción en el proceso.
Override	¿Qué condiciones justificarían apartarse de la recomendación del modelo?	Definir reglas de supervisión humana.
Riesgos percibidos	¿Qué riesgos observa antes de usar la herramienta en producción?	Identificar controles necesarios.
Mejoras futuras	¿Qué funcionalidades deberían priorizarse en una siguiente versión?	Levantar requerimientos de evolución.

Nota. Elaboración propia.

Resultados cualitativos de la validación funcional

Los resultados de la validación funcional muestran una recepción favorable del prototipo como herramienta de apoyo, no como mecanismo automático de fijación de tasas. Los usuarios destacaron tres aportes principales. Primero, la posibilidad de concentrar en una sola interfaz información que actualmente se consulta de manera dispersa. Segundo, la utilidad de comparar la recomendación contra curvas de mercado, referencias competitivas y escenarios por plazo. Tercero, el valor de contar con una explicación de drivers que permita sustentar la recomendación ante instancias internas.

La validación también confirmó que la confianza en el motor depende de su capacidad para mostrar el origen de la recomendación. Los participantes señalaron que la herramienta resulta más útil cuando no entrega únicamente una tasa final, sino cuando permite observar qué variables explican la salida, cómo se comportó el modelo en validación y qué tan cerca está la recomendación de referencias de mercado comparables.

En cuanto a las condiciones de adopción, los usuarios identificaron como prioritarios el registro de overrides, la actualización frecuente de insumos, la conservación del historial de decisiones, la explicación de la recomendación por escenario y la posibilidad de comparar tasas sugeridas contra referencias internas de fondeo. Estos hallazgos son consistentes con el diagnóstico inicial: la necesidad principal no es reemplazar el juicio experto, sino fortalecerlo mediante una capa analítica trazable, reproducible y explicable.

Finalmente, la validación permitió identificar oportunidades de mejora para versiones posteriores. Entre ellas se encuentran la incorporación de explicabilidad local mediante SHAP, la integración con fuentes intradía, la ampliación de observaciones para instrumentos IPC e IBR, la conexión con referencias FTP internas, el desarrollo de alertas de drift y la medición longitudinal del impacto del motor sobre tiempo de análisis, consistencia de decisiones y costo relativo de fondeo.

Tabla 25*Síntesis de Hallazgos de la Validación Funcional del Prototipo*

Dimensión evaluada	Hallazgo agregado	Implicación para la adopción
Coherencia de la recomendación	Los usuarios consideraron que las recomendaciones son razonables como referencia inicial, siempre que se mantenga supervisión humana.	El motor debe operar como apoyo, no como mecanismo automático.
Utilidad de benchmarks	La comparación con curvas, top bancos y referencias de mercado fue valorada como una de las salidas más útiles.	La interfaz debe priorizar la lectura comparativa antes de la tasa final.
Interpretabilidad	Los drivers del modelo facilitaron la comprensión de la recomendación, aunque se sugirió incorporar explicabilidad local en versiones futuras.	SHAP puede incorporarse como mejora posterior.
Trazabilidad	La herramienta fue percibida como útil para reconstruir decisiones y justificar tasas ex post.	Debe mantenerse historial de insumos, recomendación, usuario y override.
Uso operativo	El mayor valor se identificó antes de cotizar operaciones institucionales o validar movimientos de tasa por plazo.	El piloto debe integrarse al flujo de mesa y ALM.
Riesgos percibidos	Se identificaron riesgos de sobreconfianza, desactualización de datos y baja cobertura en IPC/IBR.	Se requieren controles de calidad, monitoreo y recalibración.

Nota. Elaboración propia.

Plan de Intervención

El diagnóstico organizacional mostró que Corficolombiana dispone de información de mercado, capacidades técnicas y una estructura institucional apta para fortalecer la decisión de pricing de CDTs; no obstante, también evidenció brechas en integración de datos, trazabilidad de la decisión, homologación de comparables y formalización del proceso. En particular, la priorización de problemas identificó como núcleo del reto la ausencia de una recomendación analítica única para pricing de CDTs, seguida por la trazabilidad parcial de la decisión comercial, la comparación no homogénea por plazo e indexador, la dependencia de criterio experto no formalizado y la falta de lineamientos explícitos de adopción y monitoreo. En consecuencia, el plan de intervención no se plantea como un desarrollo tecnológico aislado, sino como una hoja de ruta para convertir el pipeline ya construido en una capacidad institucional de apoyo a la decisión dentro de la tesorería (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011; Tabassi, 2023).

Objetivo del plan de implementación

Institucionalizar una capacidad analítica reproducible para recomendar tasas y spreads de CDTs por plazo e indexación en Corficolombiana, integrando referencias de curvas de mercado, señales macroeconómicas y variables competitivas, con trazabilidad, explicabilidad y gobierno de datos y de modelo.

Principios de implementación

La intervención se desarrollará con base en cuatro principios. El primero es la calidad, integridad y trazabilidad del dato, en línea con las buenas prácticas de agregación y gobierno de información requeridas en entornos financieros (BCBS, 2013b). El segundo es la validación, documentación y supervisión formal del modelo, de modo que la solución pueda ser utilizada con criterios de control y auditabilidad (Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, 2011). El tercero es el uso confiable y responsable de la inteligencia artificial, entendiendo que

la salida del motor prescriptivo debe estar acompañada de explicabilidad, monitoreo y revisión periódica (Tabassi, 2023). El cuarto es la apropiación por parte del usuario de negocio, de manera que la herramienta no reemplace el juicio profesional, sino que lo fortalezca con evidencia estructurada y trazable.

Articulación entre diagnóstico, pipeline y plan de intervención

El plan de intervención se deriva directamente del diagnóstico y se operacionaliza a través del pipeline analítico desarrollado en el proyecto. Las brechas priorizadas del diagnóstico definen qué debe resolverse; las fases de implementación establecen cómo debe resolverse; y el pipeline aporta con qué componentes técnicos se resuelve. Desde esta lógica, la intervención propone cerrar la fragmentación actual del proceso de pricing mediante una arquitectura que integre ingesta automatizada de insumos de mercado, homologación de referencias por plazo e indexación, construcción del basis de emisión, modelado supervisado, explicabilidad y una capa final de uso para tesorería. Así, el plan no es un capítulo separado del desarrollo técnico, sino la traducción organizacional del pipeline a una ruta de adopción real.

Fase 1. Formalización funcional y de gobierno

La primera fase busca cerrar la brecha de formalización del proceso. Su propósito es definir con claridad el alcance operativo del motor prescriptivo, los responsables de cada etapa y las reglas bajo las cuales la herramienta será utilizada dentro del proceso de pricing. En esta etapa se documentarán el propósito del modelo, las variables de entrada, los supuestos principales, los criterios de uso y los casos en los que una recomendación podrá ser ajustada o descartada por criterio experto. También se establecerán responsables de negocio, riesgo, tecnología y seguimiento, así como una matriz básica de roles y responsabilidades. El resultado esperado es una definición funcional del motor, acompañada de lineamientos de gobierno que permitan su uso bajo criterios de control, trazabilidad y supervisión humana.

Fase 2. Industrialización del pipeline de datos

La segunda fase responde principalmente a la brecha de integración de datos. En esta etapa se consolidará la adquisición, persistencia y transformación de los insumos críticos del proceso, garantizando que la información de mercado, curvas y señales macroeconómicas fluya hacia una capa analítica única y reproducible. Esto implica formalizar la descarga e ingesta de archivos Precia, estructurar el almacenamiento en formatos analíticos, versionar insumos, mantener bitácoras de ejecución y documentar reglas de transformación. También supone estabilizar componentes ya contruidos, como el maestro de TES, el staging histórico de CDTs, la curva TES continua mediante PCHIP y el dataset maestro para modelado. El resultado esperado es una base de datos trazable, reutilizable y consistente que reduzca el esfuerzo manual y sirva como soporte estable para la decisión de pricing.

En esta fase, la industrialización debe considerar el paso desde ejecución local hacia una operación administrada, idealmente en infraestructura corporativa o nube aprobada por la entidad. Esto incluye programación automática de tareas, separación de ambientes de desarrollo, pruebas y producción, control de accesos, versionamiento de datos y registro de artefactos.

Fase 3. Validación analítica y selección del modelo operativo

La tercera fase se enfoca en resolver la ausencia de una recomendación analítica única y consistente. Aquí se consolidará el modelo que se adoptará como referencia operativa dentro del proceso de pricing. La selección no debe basarse únicamente en métricas de error, sino también en criterios de estabilidad, interpretabilidad y coherencia financiera. En esta etapa se documentará la lógica del modelo, las variables de entrada, los horizontes de actualización, los criterios de recalibración y los umbrales de alerta. De igual forma, se documentará la lógica de los modelos por escenario, diferenciando el target utilizado para tasa fija, spread IBR, spread IPC y basis general. La selección del modelo operativo no debe basarse únicamente en métricas de error, sino también en estabilidad

temporal, interpretabilidad, tamaño de muestra disponible y coherencia financiera de los drivers identificados.

Fase 4. Despliegue y adopción por usuario de negocio

La cuarta fase está orientada a cerrar la brecha de adopción y uso. Un motor prescriptivo solo genera valor si puede integrarse al flujo real de tesorería y si sus usuarios entienden por qué recomienda una determinada tasa o spread. Por ello, esta fase contempla la puesta en operación de una interfaz que muestre la recomendación de pricing, su comparación frente a benchmarks de mercado, los principales drivers de la salida y un registro trazable de la decisión tomada. También implica acompañar el despliegue con sesiones de entrenamiento, apropiación funcional y validación con usuario de negocio. El resultado esperado es una operación piloto controlada en la que la herramienta se convierta en apoyo efectivo a la definición de tasas de captación, sin sustituir el juicio profesional.

El despliegue productivo de la aplicación deberá realizarse en un entorno seguro y escalable, evitando dependencia de ejecución local. La interfaz podrá publicarse mediante un servidor corporativo, contenedor Docker, plataforma cloud o servicio administrado compatible con las políticas de tecnología y seguridad de la entidad.

Fase 5. Monitoreo y mejora continua

La quinta fase se orienta a garantizar la sostenibilidad de la intervención. Su propósito es evitar que la solución quede como un prototipo estático o aislado del proceso real. Para ello se definirán indicadores de calidad del dato, desempeño del modelo, frecuencia de uso, consistencia de recomendaciones, necesidad de recalibración y revisión de deriva. También se establecerá una instancia periódica de revisión para evaluar cambios en las condiciones de mercado, evolución de la competencia y oportunidades de mejora. El resultado esperado es una capacidad analítica institucionalizada, con seguimiento continuo y posibilidades reales de ajuste en el tiempo.

Estructura del plan por componentes

Desde el punto de vista operativo, el plan de intervención se materializa en cuatro componentes que se corresponden con la arquitectura del pipeline desarrollado. El primero es el componente de adquisición y almacenamiento de datos, encargado de capturar, organizar y preservar la información de mercado y macroeconómica. El segundo es el componente de transformación y homologación, donde se construyen referencias comparables por plazo e indexación, se consolidan curvas y se calcula el basis de emisión. El tercero es el componente de modelado y evaluación, en el que se entrenan, comparan e interpretan los modelos que soportan la recomendación. El cuarto es el componente de despliegue y uso de negocio, donde la salida del motor se traduce en una herramienta útil para tesorería, con trazabilidad e historial de decisiones.

Esta estructura por componentes permite conectar de manera natural las brechas del diagnóstico con los entregables técnicos del proyecto. La integración de datos responde a la fragmentación de insumos; la homologación de comparables responde a la dificultad de comparar plazos e indexadores; el modelado responde a la ausencia de una recomendación analítica única; y la capa de interacción responde a la necesidad de explicabilidad, uso y control dentro del proceso real.

El pipeline desarrollado constituye el habilitador técnico del plan de intervención. En consecuencia, sus componentes no deben leerse como un desarrollo independiente del capítulo metodológico, sino como la infraestructura analítica que permite responder de manera concreta a las brechas priorizadas en el diagnóstico y soportar la institucionalización del motor prescriptivo dentro del proceso de tesorería.

Tabla 26

Pipeline Analítico Integral

Fuentes de datos	Ingesta y almacenamiento	Transformación y homologación	Construcción de variables y target	Modelado y evaluación	Despliegue y uso de negocio
Precia SX, SC, SB y SwapCC Series BanRep Documentos institucionales	Descarga automatizada Lectura y parseo Persistencia en Parquet Bitácoras	Maestro TES Staging CDTs Curva TES continua (PCHIP) Homologación por plazo e indexación	Cálculo de basis Variables macro Señales competitivas Dataset maestro y limpio	Torneo de modelos Selección del champion model Interpretabilidad Métricas	Motor prescriptivo App de negocio Trazabilidad de la decisión Soporte a tesorería

Nota. Elaboración propia.

Tabla 27

Trazabilidad del Dato Hasta la Recomendación

Dato de origen	Regla de transformación	Salida analítica	Uso en el modelo	Decisión de negocio
Curvas, emisiones, series macro y competencia	Limpieza, homologación, interpolación, spreads y control de leakage	Variables explicativas Target basis/spread Panel de comparables	Predicción Ranking de drivers Escenarios	Recomendación de tasa/spread Justificación trazable Registro de la decisión

Nota. Elaboración propia.

Tabla 28

Gobierno y Control del Motor Prescriptivo

Gobierno de datos	Gobierno del modelo	Adopción del usuario	Monitoreo continuo
Calidad del dato Versionamiento Diccionario Trazabilidad	Documentación Validación Overrides Segregación de funciones	Explicabilidad Capacitación Historial de decisiones Supervisión humana	KPIs de uso Drift Recalibración Revisión periódica

Nota. Elaboración propia.

Relación entre brechas priorizadas, acciones y entregables

La ausencia de una recomendación analítica única para pricing de CDTs se atiende mediante la consolidación del motor prescriptivo y la selección del modelo operativo. La trazabilidad parcial de la decisión comercial se aborda con bitácoras, versionamiento, documentación de variables y registro de decisiones. La comparación no homogénea por plazo e indexador se resuelve con la construcción de curvas comparables, la homologación técnica de instrumentos y el cálculo del basis. La dependencia elevada de criterio experto no formalizado se reduce mediante reglas documentadas, explicabilidad del modelo y uso de benchmarks homogéneos. Finalmente, la falta de lineamientos de adopción y monitoreo se atiende con gobierno de datos, gobierno de modelo, capacitación a usuarios y definición de indicadores de seguimiento.

En términos prácticos, esto significa que los scripts y artefactos técnicos del proyecto no deben verse como piezas aisladas, sino como habilitadores concretos de la intervención. Los procesos de descarga e ingesta formalizan la capa de datos; el maestro de TES, el staging de CDTs y la curva PCHIP permiten homologar comparables; el cálculo de basis y el feature engineering alimentan el dataset maestro; el modelado y la interpretabilidad dan sustento a la recomendación; y la aplicación de negocio traduce el desarrollo analítico en una herramienta usable dentro del proceso de tesorería.

Con el fin de mostrar de manera explícita cómo el plan responde a los hallazgos del diagnóstico, a continuación, se presenta el mapeo entre las brechas priorizadas, los componentes técnicos que las atienden y los entregables que permitirían materializar la intervención en el entorno organizacional.

Tabla 29

Mapeo Entre Brechas Priorizadas, Componentes Técnicos y Entregables de Intervención

Brecha priorizada	Acción de intervención	Componentes / scripts asociados	Entregable técnico	Entregable institucional
-------------------	------------------------	---------------------------------	--------------------	--------------------------

Ausencia de una recomendación analítica única para pricing de CDTs	Consolidar modelos por escenario de pricing: tasa fija, spread IBR, spread IPC y basis general	6_Modelado_ML_Avanzado.py; modelos .joblib; metadatos por escenario	Modelos ganadores por escenario y ranking de desempeño	Criterio formal de uso del modelo por tipo de CDT
Trazabilidad parcial de la decisión comercial	Registrar insumos, transformaciones, modelos usados, métricas y drivers de recomendación	Manifest de descargas; Parquets intermedios; 7_Interpretabilidad.py; 8_App_Streamlit.py	Bitácoras, datasets versionados, artefactos de interpretabilidad y predicciones	Evidencia auditable para justificar decisiones de tasa o spread
Comparación no homogénea por plazo e indexador	Construir referencias comparables mediante curvas TES, curvas AAA, swap, PCHIP y variables por plazo	2.1_Maestro_TESSX.py; 3.2_Curva_TES_Pchip.py; 4.1_Maestro_CceroC.py; 4.2_Feature_Engineering_Macro.py	Curvas interpoladas, variables soberanas, AAA, swap y competitivas	Política funcional de comparación por plazo e indexador
Dependencia elevada de criterio experto no formalizado	Complementar el juicio experto con modelos explicables, benchmarks y reglas documentadas	5_Preprocesamiento_ML.py; 6_Modelado_ML_Avanzado.py; 7_Interpretabilidad.py	Dataset limpio, modelos validados, importancia de variables y diagnósticos	Guía de uso del motor como apoyo, no sustituto del criterio profesional
Fragmentación de datos de mercado y macroeconómicos	Industrializar la ingesta y persistencia de fuentes Precia y Banco de la República	1.2_Precia.py; 1.1_Banrep.py; 2.2_Maestro_CDT SX.py; 4.1_Maestro_CceroC.py	Data lake en Parquet, maestros TES/CDT, curvas y series macro	Proceso operativo de actualización con responsables definidos
Riesgo de fuga de información y baja reproducibilidad	Mantener reglas explícitas de exclusión, deduplicación, validación temporal y control de leakage	5_Preprocesamiento_ML.py; 6_Modelado_ML_Avanzado.py	Dataset limpio, split temporal, TimeSeriesSplit, variables excluidas documentadas	Documento de validación metodológica del modelo

Baja adopción por usuario de negocio	Desplegar una interfaz que muestre datos, torneo, interpretabilidad y simulador de pricing	8_App_Streamlit.py	Aplicación Streamlit con pestañas de datos, torneo, interpretabilidad y motor de pricing	Prototipo funcional para piloto controlado con Tesorería / ALM
Falta de monitoreo y gobierno posterior	Definir controles de desempeño, calidad de datos, recalibración y supervisión humana	Componentes transversales del pipeline y plan de intervención	Métricas, reportes, logs y criterios de recalibración	Comité o instancia funcional de seguimiento del motor
Dependencia de ejecución local y baja escalabilidad	Migrar el prototipo hacia una arquitectura administrada con orquestación, registro de modelos y monitoreo	8_App_Streamlit.py, pipeline de datos, repositorio de modelos, scheduler u orquestador	Arquitectura objetivo MLOps y despliegue seguro de la app	Ruta de transición de prototipo local a solución institucional

Nota. Elaboración propia.

Indicadores de éxito

Los indicadores mínimos de éxito del plan incluyen la reducción del esfuerzo manual en la consolidación de insumos, la mejora en la capacidad de justificar ex post una decisión de pricing, la existencia de una referencia analítica única para comparar alternativas de tasa o spread, el uso recurrente de la herramienta por parte de tesorería y la implementación de controles formales de gobierno de datos y de modelo. Estos indicadores no solo permiten evaluar si el desarrollo técnico funciona, sino si la intervención realmente logra fortalecer la calidad de la decisión comercial y la trazabilidad del proceso.

Resultado esperado de la intervención

La intervención no pretende reemplazar la decisión humana, sino mejorar la forma en que esta se soporta. Su resultado esperado es un proceso de pricing de CDTs más consistente, trazable, comparable frente al mercado y gobernable desde la perspectiva de negocio, riesgo y tecnología. En este sentido, el valor del plan no está únicamente en construir un pipeline o entrenar un modelo, sino en convertir ese desarrollo en una capacidad institucional útil para el fondeo, coherente con las brechas identificadas en el diagnóstico y sostenible en el tiempo.

Con el fin de llevar la propuesta desde una arquitectura analítica viable hacia una capacidad institucional implementable, el plan de intervención se complementa con una ruta operativa de ejecución. Esta ruta incorpora temporalidad, responsables, recursos mínimos y criterios de seguimiento, de manera que la propuesta no se limite a un desarrollo técnico, sino que se traduzca en una intervención empresarial gestionable dentro del entorno de tesorería de Corficolombiana. Esta aproximación es consistente con la formulación de planes de mejora en trabajos dirigidos, en los que la viabilidad de la propuesta depende no solo de su rigor técnico, sino también de su capacidad de implementación y sostenibilidad organizacional (Project Management Institute, 2021)

Cronograma propuesto de implementación

La implementación de la propuesta puede desarrollarse en un horizonte de seis meses, organizado en fases secuenciales con algunos frentes parcialmente superpuestos. Este horizonte es razonable para una adopción piloto, ya que permite formalizar el gobierno del motor, estabilizar la capa de datos, validar el modelo operativo, desplegar la herramienta a usuario de negocio y establecer un esquema inicial de monitoreo.

Tabla 30

Cronograma Propuesto de Implementación del Motor Prescriptivo de Pricing de CDTs

Fase	Actividad principal	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6
1	Formalización funcional y de gobierno	X	X				
2	Industrialización del pipeline de datos	X	X	X			
3	Validación analítica y selección del modelo operativo		X	X	X		
4	Despliegue y adopción por usuario de negocio				X	X	
5	Monitoreo, ajustes y mejora continua inicial					X	X

Nota. Elaboración propia.

Responsables propuestos

La adopción del motor prescriptivo requiere una lógica de responsabilidad compartida.

Tesorería debe liderar el uso funcional de la herramienta; Riesgo y Cumplimiento deben acompañar la validación y el marco de control; Tecnología o soporte analítico deben velar por la estabilidad técnica del pipeline y la aplicación; y la alta dirección funcional debe asegurar patrocinio, priorización y seguimiento. Esta distribución de responsabilidades es coherente con las buenas prácticas de gobierno de modelo y con la necesidad de evitar que una solución analítica de impacto financiero quede concentrada en una sola persona o área.

Tabla 31

Responsables Propuestos Para la Implementación

Frente	Responsable principal	Apoyo requerido	Rol esperado
Definición funcional del motor	Vicepresidencia de Tesorería	ALM, Mesa Moneda Legal, Mesa Institucional	Definir criterios de uso, salidas requeridas y condiciones de adopción
Gobierno de datos y modelo	Gobierno, Riesgo y Cumplimiento	Tesorería, analítica, tecnología	Validar supuestos, controles, documentación y seguimiento

Pipeline de datos	Equipo analítico / soporte técnico	Tesorería, tecnología	Mantener ingesta, calidad de datos, versionamiento y bitácoras
Validación del modelo	Equipo analítico y Riesgo	Tesorería	Revisar desempeño, estabilidad e interpretabilidad
Despliegue y uso de negocio	Tesorería	Tecnología, analítica	Operar la herramienta, registrar decisiones y retroalimentar mejoras
Monitoreo periódico	Comité funcional de seguimiento	Tesorería, Riesgo, Tecnología	Revisar KPIs, drift, recalibración y necesidades de ajuste

Nota. Elaboración propia.

Recursos mínimos requeridos

La implementación de la propuesta no requiere, en principio, una infraestructura completamente nueva, ya que el proyecto ya demostró viabilidad técnica mediante el pipeline desarrollado. Sin embargo, sí exige recursos mínimos para su institucionalización: tiempo de usuarios expertos, soporte analítico, capacidades de almacenamiento y procesamiento, y espacios de seguimiento funcional. En este sentido, la intervención debe entenderse más como un proceso de formalización, adopción y gobierno que como una adquisición tecnológica de gran escala.

Tabla 32

Recursos Mínimos Requeridos Para la Implementación

Tipo de recurso	Requerimiento	Propósito
Recurso humano	Tiempo de Tesorería, ALM, Riesgo y soporte analítico	Validación funcional, ajuste del modelo y adopción
Recurso tecnológico	Entorno corporativo o cloud para ejecución programada del pipeline, almacenamiento versionado, registro de modelos y despliegue seguro de la app	Operación escalable, trazable y no dependiente de ejecución local
Recurso de gestión	Espacios periódicos de seguimiento	Monitoreo, revisión de KPIs y toma de decisiones

Recurso documental	Manual funcional, documentación técnica, bitácoras	Trazabilidad, capacitación y gobierno
Recurso de control	Definición de comité o instancia de revisión	Supervisión del modelo y mejora continua

Nota. Elaboración propia.

Riesgos de implementación y medidas de mitigación

Como toda intervención de base analítica, la propuesta enfrenta riesgos de implementación que deben ser gestionados desde el inicio. Entre ellos se destacan la baja apropiación por parte de usuario de negocio, la dependencia excesiva de unos pocos actores técnicos, posibles problemas de calidad o actualización de datos, y el riesgo de sobreconfianza en la salida del modelo. Por ello, la puesta en marcha del motor prescriptivo debe acompañarse de capacitación, documentación, validación periódica, monitoreo de drift y mantenimiento del principio de supervisión humana sobre la decisión final (Tabassi, 2023)

Tabla 33
Riesgos de Implementación y Mitigación

Riesgo	Posible efecto	Medida de mitigación
Baja adopción por usuario de negocio	Subutilización del motor	Entrenamiento, explicabilidad y validación con tesorería
Calidad insuficiente del dato	Recomendaciones inconsistentes	Controles de calidad, bitácoras y versionamiento
Sobreconfianza en la salida del modelo	Uso acrítico de la recomendación	Política de supervisión humana y registro de overrides
Dependencia de pocos actores técnicos	Riesgo operativo y de continuidad	Documentación, transferencia de conocimiento y manuales
Drift por cambios de mercado	Pérdida de desempeño	Monitoreo periódico y recalibración

Dependencia de ejecución local	Riesgo de continuidad, baja escalabilidad y dificultad de auditoría	Despliegue en infraestructura administrada, contenedorización, orquestación del pipeline y control de versiones
--------------------------------	---	---

Nota. Elaboración propia.

Conclusiones

El marco teórico construido permitió identificar los conceptos fundamentales que articulan el problema de pricing de CDTs con el campo de la ciencia de datos. La decisión de pricing no es un acto comercial aislado, sino una función que conecta la estructura temporal de tasas soberanas, la curva swap, las señales macroeconómicas del Banco de la República y la presión competitiva del sistema financiero. La noción de basis de emisión —derivada de la literatura sobre curvas de referencia y spreads de crédito— se identificó como el concepto que permite convertir precios absolutos de captación en una medida relativa y comparable entre instrumentos de distinto plazo e indexador. El marco normativo colombiano, en particular la Circular Externa 025 de 2022 de la SFC sobre RTILB y los estándares BCBS 239, SR 11-7 y NIST AI RMF 1.0, delimitó las exigencias de trazabilidad, gobierno de datos y gobierno de modelo que debe cumplir cualquier solución analítica aplicada a un proceso de decisión financiera sensible.

El diagnóstico organizacional —construido a partir del análisis estratégico del entorno (Porter y PESTEL), el levantamiento AS-IS del proceso y el instrumento aplicado a cinco informantes clave— confirmó que la principal brecha de Corficolombiana no es la falta de información de mercado, sino la ausencia de una capa analítica integrada que convierta esas señales en decisiones trazables y reproducibles. Todos los informantes reportaron que consultan curvas TES, swap, referencias AAA y señales de competencia de manera sistemática, pero que la trazabilidad de la decisión es limitada y la dependencia de criterio experto individual es máxima. El instrumento cuantificó que solo uno de cinco informantes considera que la justificación de la tasa queda documentada de forma verificable "Siempre", mientras que la dependencia de personas específicas fue calificada como "Siempre" por los cinco participantes. Este hallazgo valida la pertinencia del motor prescriptivo no como sustituto del criterio experto, sino como infraestructura que lo institucionaliza y lo hace auditable.

Se diseñó e implementó una arquitectura de trece componentes de software que transforma datos crudos de Precia y del Banco de la República en un dataset analítico listo para modelado. El pipeline implementa un proceso incremental que evita reprocesar el histórico completo, genera bitácoras de ejecución, persiste resultados intermedios en formato Parquet y construye la curva soberana continua mediante interpolación PCHIP para todos los plazos de 1 a 5.475 días. El pipeline permitió construir una base analítica diferenciada por escenarios de pricing. En lugar de tratar el pricing de CDTs como un problema de target único, la solución separó la modelación en tasa fija, spread IBR, spread IPC y basis general. Esta decisión metodológica fue fundamental para respetar la naturaleza financiera de cada modalidad de captación y mejorar la coherencia entre el target, las variables explicativas y la recomendación final.

Se desarrolló un torneo de modelos supervisados por escenario de pricing, evaluando Ridge, ElasticNet, Random Forest, XGBoost, LightGBM y Stacking. Los resultados muestran que no existe un único modelo dominante para todos los casos. En tasa fija, ElasticNet presentó el mejor desempeño; en spread IBR, Ridge fue el modelo más estable; en spread IPC, ElasticNet obtuvo el mejor resultado aunque con una muestra más reducida; y en basis general, XGBoost capturó mejor las relaciones no lineales entre variables competitivas, plazo, curvas y señales de mercado. Esta diferenciación por escenario fortalece la coherencia financiera del motor, dado que cada modalidad de CDT responde a una lógica distinta de formación de tasas y spreads.

El plan de intervención propuesto organiza la adopción del motor en cinco fases —formalización de gobierno, industrialización del pipeline, validación analítica, despliegue con usuario de negocio y monitoreo continuo— en un horizonte de seis meses. La validación funcional con usuarios de mesa confirmó que los criterios de adopción identificados en el diagnóstico —explicabilidad, baja latencia, override trazable, historial de decisiones y rol de apoyo al criterio experto— son consistentes con las expectativas de uso del prototipo. La valoración unánime del proyecto como "Muy alta" por parte de los

tres informantes, junto con su disposición a participar en validaciones posteriores, constituye evidencia práctica de que la propuesta responde a una necesidad real y reconocida dentro del proceso de tesorería.

La investigación demuestra que el pricing de CDTs en Corficolombiana es un problema abordable desde una intervención de ciencia de datos aplicada, y que el valor del motor prescriptivo no reside únicamente en su desempeño técnico estadístico, sino en su capacidad de institucionalizar una forma más disciplinada, trazable y explicable de decidir sobre el fondeo. La viabilidad de largo plazo de la propuesta depende de que su adopción esté acompañada de gobierno de datos, validación periódica, supervisión humana y actualización continua ante cambios en las condiciones del mercado.

Todos los resultados técnicos reportados corresponden a la ejecución del pipeline con corte al 03 de mayo de 2026; por tanto, las métricas y salidas del modelo deben interpretarse como una fotografía reproducible de ese momento de información y no como valores estáticos e invariables del sistema.

Recomendaciones

Se recomienda que Corficolombiana adopte el motor prescriptivo como herramienta de apoyo a la decisión y no como mecanismo automático de fijación de tasas. La supervisión humana debe mantenerse como principio central del proceso, especialmente en escenarios atípicos de mercado, eventos de estrés de liquidez o decisiones comerciales estratégicas de alto volumen.

Se recomienda formalizar un esquema de gobierno de datos y gobierno del modelo para el proceso de pricing de CDTs. Esto implica designar responsables formales por frente (tesorería, riesgo, tecnología y analítica), definir políticas de actualización del pipeline, establecer controles de calidad del dato, documentar criterios de recalibración del modelo y crear una instancia periódica de revisión de desempeño y deriva.

Se recomienda fortalecer la integración de la herramienta con el flujo real de tesorería, priorizando las salidas de mayor utilidad para el usuario de mesa: recomendación por plazo e indexador, comparación frente a benchmarks de mercado actualizados, descomposición de drivers de la recomendación y registro trazable de la decisión tomada —incluyendo cualquier ajuste respecto a la sugerencia del modelo y su justificación.

Se recomienda planificar la recalibración del modelo al menos cada seis meses, o cuando las condiciones de política monetaria cambien de manera sustancial, dado que el período de entrenamiento del modelo actual corresponde principalmente a un ciclo de tasas altas (2022-2024) cuyo comportamiento puede diferir del ciclo de normalización en curso.

Se recomienda formalizar una etapa periódica de validación con usuarios de mesa, en la cual se comparen las recomendaciones del motor frente a decisiones reales o simuladas de pricing. Esta validación debe registrar si la recomendación fue aceptada, ajustada o descartada, así como el motivo

del override. Este mecanismo permitiría convertir el uso cotidiano de la herramienta en una fuente de aprendizaje supervisado y en evidencia de gobierno del modelo.

Se recomienda evolucionar el prototipo local hacia una arquitectura MLOps institucional, soportada en infraestructura corporativa o nube aprobada por la entidad. Esta evolución debe incluir orquestación automática del pipeline, versionamiento de datos y modelos, registro de experimentos, monitoreo de drift, control de accesos, despliegue seguro de la aplicación y trazabilidad de predicciones y overrides. De esta manera, el motor podrá pasar de una prueba funcional a una capacidad operativa sostenible y auditable.

Por último, se recomienda que futuras extensiones del proyecto profundicen en dos frentes: la incorporación de valores SHAP para la explicabilidad local de cada recomendación individual —lo que permitiría al usuario ver qué factores específicos explican la tasa sugerida para una emisión concreta—, y la evaluación longitudinal del impacto del motor sobre la consistencia del pricing, el costo relativo de fondeo y la reducción del esfuerzo manual de la mesa de tesorería.

Referencias

Alfaro, R. (2009). *LA CURVA DE RENDIMIENTO BAJO NELSON-SIEGEL*.

<http://www.bcentral.cl/esp/estpub/estudios/dtbc.ExistelaPosibilidaddeSolicitarUnaCopiaImpresada>

[tp://www.bcentral.cl/eng/stdpub/studies/workingpaper](http://www.bcentral.cl/eng/stdpub/studies/workingpaper).

Andrada Félix, J., Fernández Pérez, A., & Fernández Rodríguez, F. (2014). La estructura temporal de los tipos de interés: estrategias de negociación en renta fija. *Cuadernos de Economía*, 37(105), 131–149. <https://doi.org/10.1016/j.cesjef.2013.09.001>

Apache Software Foundation. (2024). *Parquet Documentation*. <https://parquet.apache.org/docs/>

Asobancaria. (2020). *Edición 1253 Gestión del riesgo de tasa de interés del libro bancario (IRRBB): regulación, avances y retos*. <https://asobancaria.com/ws/semanas-economicas/1253-BE.pdf>

Banco de la República. (s/f-a). *Certificado de depósito a término (CDT)*. Recuperado el 15 de abril de 2026, de <https://www.banrep.gov.co/es/glosario/certificado-deposito-termino-cdt>

Banco de la República. (s/f-b). *Indicador Bancario de Referencia (IBR)*. Recuperado el 15 de abril de 2026, de https://suameca.banrep.gov.co/estadisticas-economicas/informacionSerie/241/tasas_interes_indicador_bancario_referencia_ibr

Banco de la República. (s/f-c). *Tasa de cambio o tasa de cambio representativa del mercado (TRM)*. Recuperado el 15 de abril de 2026, de <https://www.banrep.gov.co/es/glosario/tasa-cambio-trm>

Banco de la República. (s/f-d). *Tasa de interés de intervención o de política monetaria*. Recuperado el 15 de abril de 2026, de <https://www.banrep.gov.co/es/glosario/tasa-interes-politica-monetaria>

Banco de la República. (s/f-e). *Unidad de valor real (UVR)*. Recuperado el 15 de abril de 2026, de <https://www.banrep.gov.co/es/glosario/uvr>

Banco de la República. (2024a). *Reporte Estabilidad Financiera Primer Semestre 2024*.

Banco de la República. (2024b). *Reporte Estabilidad Financiera Segundo Semestre 2024*.

Banco de la República. (2025). *PIB trimestral observado - crecimiento anual serie original*.

https://suameca.banrep.gov.co/estadisticas-economicas/informacionSerie/500011/producto_interno_bruto_pib_trimestral_observado_crecimiento_anual_serie_original

BCBS. (2013a). *Basel Committee on Banking Supervision Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools*. https://www.bis.org/basel_framework/

BCBS. (2013b). *Basel Committee on Banking Supervision Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools*. <https://www.bis.org/publ/bcbs238.pdf>

BCBS. (2013c). *Basel Committee on Banking Supervision Principles for effective risk data aggregation and risk reporting*. www.bis.org

BCBS. (2016). *Basel Committee on Banking Supervision Standards Interest rate risk in the banking book*. https://www.bis.org/basel_framework/

Belltech. (2023, febrero 2). *Potenciando el Negocio de las Mesas de Dinero*. <https://belltech.la/blog/potenciado-el-negocio-de-las-mesas-de-dinero/>

Bernal Torres, C. A. (2016). *Metodología de la investigación* (PEARSON, Ed.; 4a ed.).

BIS. (2005). *Zero-coupon yield curves: technical documentation*.

BIS. (2015). *Interest rate risk in the banking book*. <https://www.bis.org/bcbs/publ/d319.pdf>

Board of Governors of the Federal Reserve System, & Office of the Comptroller of the Currency. (2011). *SUPERVISORY GUIDANCE ON MODEL RISK MANAGEMENT CONTENTS*.

Breiman, L. (2001). *Random Forests* (Vol. 45).

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K.

(2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing: JRN*, 25(8), 652. <https://doi.org/10.1177/1744987120927206>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *Step-by-step data mining guide*.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>

Corficolombiana. (s/f-a). *Nuestra Historia*. Recuperado el 23 de septiembre de 2025, de <https://www.corfi.com/es/nuestra-historia>

Corficolombiana. (s/f-b). *Portafolio de Productos de Corficolombiana*. Recuperado el 9 de septiembre de 2025, de <https://www.corfi.com/documents/20123/107225/Informaci%C3%B3n%20de%20Productos%20y%20servicios.pdf>

Corficolombiana. (s/f-c). *Quiénes somos Corficolombiana*. Recuperado el 23 de septiembre de 2025, de <https://www.corfi.com/es/quienes-somos>

Corficolombiana. (2025a). *Estados Financieros Separados al 31 de Diciembre de 2024*.

Corficolombiana. (2025b). *INFORME DE GESTIÓN CORFISOSTENIBLE 2024*.

<https://www.corfi.com/documents/20123/1487206/informe-de-gestion-corfisostenible-2024.pdf>

Corficolombiana. (2025c). *Organigrama General*.

<https://www.corfi.com/documents/20123/0/organigrama-general-corfi.pdf>

Corficolombiana. (2025d). *Resultados financieros - 1T 2025*. 3–4.

<https://www.corfi.com/documents/20123/0/reporte-corfi-1t-25.pdf>

DAMA International. (2017). *Data Management Body of Knowledge*.

DANE. (2025a). *Boletín técnico Índice de Precios al consumidor (IPC)*.

<https://www.dane.gov.co/files/operaciones/IPC/dic2024/bol-IPC-dic2024.pdf>

DANE. (2025b). *Boletín técnico Producto Interno Bruto (PIB)*.

<https://www.dane.gov.co/files/operaciones/PIB/bol-PIB-ltrim2025.pdf>

Findeter. (2024, junio 14). *¿Cómo va el sector financiero y cuáles son sus principales retos?*

<https://www.findeter.gov.co/blog/informes-economicos/como-va-el-sector-financiero-y-cuales-son-sus-principales-retos>

FINNOSUMMIT. (2024). *2 de cada 3 Fintech en Colombia implementan inteligencia artificial para sus operaciones - Finnosummit*. <https://www.finnosummit.com/radar/fintech-radar-colombia-2024/>

Fitzpatrick, R. (2013). *The Mom Test*.

Friedman, J. H. (2001). 999 REITZ LECTURE GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE 1. En *The Annals of Statistics* (Vol. 29, Número 5).

FSB. (2024). *The Financial Stability Implications of Artificial Intelligence*.

Google Cloud. (2024). *MLOps: flujos de procesamiento de entrega continua y automatización en el aprendizaje automático | Cloud Architecture Center | Google Cloud Documentation*.

<https://docs.cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning?>

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction*.

Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la Investigación: Las Rutas Cuantitativa, Cualitativa y Mixta* (S. A. McGRAW-HILL INTERAMERICANA EDITORES, Ed.; 1a ed.).

Herrera, D., Pereira, W., Volochen, L., María, A., & Moreno, Z. (2023). Las finanzas abiertas en América Latina y el Caribe: grandes oportunidades, grandes desafíos. *Open Finance in Latin America and the Caribbean: Great Opportunities, Large Challenges*. <https://doi.org/10.18235/0004937>

Holland & Knight. (2024, marzo 7). *La SFC expide regulación sobre las finanzas abiertas en Colombia | Insights*. <https://www.hklaw.com/en/insights/publications/2024/03/la-sfc-expide-regulacion-sobre-las-finanzas-abiertas-en-colombia>

IBM. (s/f). *Modelador IBM SPSS - CRISP-DM - Documentación de IBM*. Recuperado el 25 de marzo de 2026, de <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning with Applications in R Second Edition*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 30. <https://github.com/Microsoft/LightGBM>.

KPMG. (2016). *Fund Transfer Pricing*. <https://assets.kpmg.com/content/dam/kpmg/rs/pdf/2018/01/rs-Fund-Transfer-Pricing.pdf>

Le Floc', F. (2013). *Stable Interpolation for the Yield Curve*.

<https://doi.org/10.1080/1350486X.20xx.CATSid>

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*.

<https://github.com/slundberg/shap>

Molnar, C. (2019). *Interpretable Machine Learning A Guide for Making Black Box Models Explainable*.

<http://leanpub.com/interpretable-machine-learning>

Moodys. (2021, junio 30). *Do funds transfer pricing methodologies still work with excess deposits?*

<https://www.moodys.com/web/en/us/insights/banking/funds-transfer-pricing-methodologies.html>

Penfield, R. D., & Giacobbi, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213–225.

https://doi.org/10.1207/S15327841MPEE0804_3

Porter, M. E. (2008). *The Five Competitive Forces That Shape Strategy*. <https://hbr.org/2008/01/the-five-competitive-forces-that-shape-strategy>

Precia. (s/f). *Quiénes Somos*. Recuperado el 15 de abril de 2026, de

<https://www.precia.co/index.php/quienes-somos/>

Project Management Institute. (2021). *PMBOK Guide*. <https://lccn.loc.gov/2021011107>

Rankings Latam. (2025, mayo 13). *Competitive Performance and Financial Results of Banks and Financial*.

<https://rankingslatam.com/en-co/blogs/industry-news/competitive-performance-and-financial-results-of-banks-and-financial-institutions-in-colombia-december-2024-rankings>

- SFC. (s/f). *Certificado de Depósito a Término CDT*. Recuperado el 25 de marzo de 2026, de <https://www.superfinanciera.gov.co/preguntas-frecuentes/8/8-certificado-de-deposito-a-termino-cdt/>
- SFC. (2006). *RESOLUCIÓN NÚMERO 01822 DE 2006*.
- SFC. (2022a). *CAPÍTULO XXXI Sistema Integral de Administración de Riesgos (SIAR)*. <https://www.superfinanciera.gov.co/publicaciones/20149/normativanormativa-generalcirculares-externas-cartas-circulares-y-resoluciones-desde-el-ano-20149/>
- SFC. (2022b). *Circular Externa 025 de 2022*. <https://www.cerlatam.com/normatividad/superfinanciera-circular-externa-025-de-2022/>
- SFC. (2024). *Circular Externa 004 de 2024*.
- Streamlit. (s/f). *Deploy Streamlit apps*. Recuperado el 4 de mayo de 2026, de <https://docs.streamlit.io/deploy>
- Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. <https://doi.org/10.6028/NIST.AI.100-1>
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. En *J. R. Statist. Soc. B* (Vol. 58, Número 1). <https://academic.oup.com/jrsssb/article/58/1/267/7027929>
- Uribe Macías, M. E. (2021). *Administración estratégica. Modelo de aplicación para organizaciones latinoamericanas* (1a ed.). Ediciones de la U.
- van Dyk, A. (2025). Non-Maturing Deposits: Predictive Modelling and Risk Management. *Journal of Risk and Financial Management* 2025, Vol. 18, Page 84, 18(2), 84. <https://doi.org/10.3390/JRFM18020084>

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.

[https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. En *J. R. Statist. Soc.*

B (Vol. 67, Número 2). <https://academic.oup.com/jrsssb/article/67/2/301/7109482>

Anexos

Anexo A – Carta de Aval de la Empresa Para Realizar la Intervención



FORMATO_DE_AUTO
RIZACION_PARA_LA_IN

Bogotá, 14 de noviembre de 2025

Señores,
Subproceso de Trabajos de Grado
Universidad Ean
Ciudad

Respetados señores,

Por medio de la presente nos permitimos autorizar a Anggie Tatiana Moreno Abella y Carlos Andrés Aranda Barreto, identificados con cédula de ciudadanía 1.033.759.919 de Bogotá y 1.018.433.847 respectivamente, estudiantes del programa de Maestría en Ciencia de Datos de la Universidad Ean, para que realice en nuestra organización Corficolombiana S.A., su trabajo de grado titulado: Optimización de la Medición de Rentabilidad y Asignación de Costos Mediante Ciencia de Datos en Corficolombiana, bajo la modalidad de trabajo dirigido.

Como empresa nos comprometemos a brindar la información requerida para adelantar este proyecto académico, así mismo, autorizamos que el documento resultado de este trabajo sea publicado en el repositorio documental Minerva de la Universidad Ean.

A continuación, relacionamos los datos de la persona que será el contacto designado por la empresa.

NOMBRE DEL CONTACTO: Humberto Manuel Cerpa Jaraba
CARGO QUE OCUPA: Gerente Core de Tesorería
TELÉFONO: 3004773701
CORREO ELECTRÓNICO: humberto.cerpa@corfi.con

Cordialmente,

Humberto Cerpa J.

Humberto Manuel Cerpa Jaraba
Gerente Core de Tesorería

	Documento No. ba9d0cd4-d7a4-4c73-9ab2-2c55cd32f3a4	
	Creado el: 11/11/2025 11:29 a. m.	

Este documento es la representación de un documento original en formato electrónico. Para validar el estado actual del documento ingrese a [corficolombiana.com.co](#) y escanee el código QR.

Este documento está firmado electrónicamente, de conformidad con los estándares internacionales de firma en tanto es un documento auténtico, íntegro y disponible para consulta en línea. [i Página](#)



Nombre(s): Humberto Manuel
Apellido(s): Cerpa Jaraba
CC: 72286563
Firmó el: 11/11/2025 11:29 a. m.

Anexo B - Instrumento de medición final aplicado a informantes clave

Proceso actual de pricing de CDTs en Corficolombiana

Propósito. Este instrumento apoya el diagnóstico interno del proyecto “Motor prescriptivo de pricing de CDTs basado en curvas de mercado y aprendizaje automático: caso Corficolombiana”. Busca recoger hechos recientes, prácticas operativas, fricciones del proceso actual y criterios de adopción de una solución analítica.

Modalidad sugerida. Aplicación asistida por entrevista o diligenciamiento acompañado.

Duración estimada: 20 a 30 minutos.

Confidencialidad. La información se utilizará con fines académicos y de mejora del proceso. Las respuestas serán anonimizadas y reportadas de manera agregada.

<i>Cargo / rol</i>	<i>Área</i>
<i>Años de experiencia total</i>	<i>Años en rol actual</i>
<i>Frecuencia de participación en pricing de CDTs</i>	<i>Modalidad</i>

Consentimiento informado

He leído el propósito del instrumento, entiendo que mi participación es voluntaria y autorizo el uso académico y agregado de mis respuestas.

☐ Sí, autorizo mi participación

☐ No autorizo mi participación

A. Información general y contexto de participación

1. Describa brevemente su rol en relación con la definición, validación, aprobación o uso de tasas/spreads de CDTs.

2. ¿Con qué frecuencia participa hoy en decisiones de pricing de CDTs? (Diaria, semanal, mensual, por solicitud puntual u otra).

3. ¿Qué plazos, indexadores o tipos de negociación conoce o gestiona con mayor frecuencia?

B. Experiencias recientes del proceso actual

4. Piense en la última vez que participó en la definición o ajuste de una tasa o spread de CDT. ¿Qué ocurrió? Indique, si puede, plazo, indexador, contexto comercial o de mercado y quiénes intervinieron.

5. En ese caso reciente, ¿qué información consultó realmente antes de tomar o recomendar la decisión? Mencione fuentes concretas (por ejemplo: curva TES, curva swap, referencia AAA, competencia, liquidez interna, tasas BanRep, histórico propio, criterio comercial u otras).

6. ¿Qué parte de ese caso tomó más tiempo o generó mayor discusión? Explique por qué.

7. Recuerde un caso reciente en el que la tasa inicialmente propuesta tuvo que cambiar. ¿Qué detonó el cambio: competencia, curva de mercado, ¿cliente, liquidez, aprobación interna u otro factor?

8. Recuerde un caso reciente en el que la decisión de pricing generó duda, retrabajo, escalamiento o inconformidad posterior. ¿Qué pasó y qué lección dejó?

C. Estandarización, trazabilidad y dependencia de criterio experto

Marque la frecuencia con la que, en su experiencia reciente, ocurren los siguientes hechos en el proceso actual.

Ítem	Nunca	Rara vez	A veces	Frecuente	Siempre
9. Se consulta de forma explícita una referencia de mercado comparable antes de cotizar una tasa.	☐	☐	☐	☐	☐
10. La competencia observada influye en la decisión final de pricing.	☐	☐	☐	☐	☐
11. La justificación de la tasa queda documentada de manera verificable.	☐	☐	☐	☐	☐
12. Se requiere escalar o revisar la decisión por falta de información integrada.	☐	☐	☐	☐	☐
13. La decisión depende en alto grado del criterio o experiencia de personas específicas.	☐	☐	☐	☐	☐

14. ¿Qué partes del proceso actual dependen más del criterio experto individual y menos de reglas o información trazable? Dé ejemplos concretos.

15. Si un tercero le pidiera hoy justificar por qué se ofreció una tasa específica hace una o dos semanas, ¿qué tan fácil sería reconstruir esa decisión y con qué soportes contaría?

D. Uso real de información y necesidades de datos

Indique con qué frecuencia usa hoy cada una de las siguientes fuentes o señales al participar en decisiones de pricing de CDTs.

Ítem	Nunca	Baja	Media	Alta	Muy alta
16. Curva TES / referencia soberana	☐	☐	☐	☐	☐
17. Curva swap o monetaria	☐	☐	☐	☐	☐
18. Referencia AAA o corporativa comparable	☐	☐	☐	☐	☐
19. Tasas BanRep / señales macro	☐	☐	☐	☐	☐
20. Cotizaciones o señales de competencia	☐	☐	☐	☐	☐

21. <i>Histórico de decisiones propias</i>	☐	☐	☐	☐	☐
22. <i>Señales de liquidez interna o costo de fondeo</i>	☐	☐	☐	☐	☐

23. ¿Qué información, reporte o visualización le habría ayudado en un caso reciente y hoy no se encuentra integrada o disponible oportunamente?

24. Si pudiera priorizar tres variables o salidas que una herramienta analítica debería mostrarle siempre antes de decidir una tasa, ¿cuáles serían y por qué?

E. Criterios de adopción del motor prescriptivo

25. ¿En qué momento exacto del proceso actual le sería más útil una herramienta que sugiera tasas o spreads con soporte de curvas, variables macroeconómicas y competencia?

26. ¿Qué condiciones mínimas tendría que cumplir esa herramienta para que usted confíe en ella? Considere, por ejemplo, trazabilidad, explicabilidad, actualización, escenarios, historial o integración al flujo actual.

Valore la utilidad potencial de los siguientes entregables de una herramienta de pricing.

<i>Ítem</i>	<i>Muy baja</i>	<i>Baja</i>	<i>Media</i>	<i>Alta</i>	<i>Muy alta</i>
27. <i>Recomendación de tasa/spread por plazo e indexador</i>	☐	☐	☐	☐	☐
28. <i>Comparación frente a referencias de mercado y competencia</i>	☐	☐	☐	☐	☐
29. <i>Explicación de los principales drivers de la recomendación</i>	☐	☐	☐	☐	☐
30. <i>Simulación de escenarios de curva, competencia o liquidez</i>	☐	☐	☐	☐	☐
31. <i>Registro trazable de la decisión y sus soportes</i>	☐	☐	☐	☐	☐
32. ¿Qué riesgos o reservas le genera el uso de una solución prescriptiva en este proceso?					

33. En términos generales, ¿cómo valora la utilidad de este proyecto para fortalecer el pricing de CDTs en Corficolombiana?

☐ 1 Muy baja ☐ 2 Baja ☐ 3 Media ☐ 4 Alta ☐ 5 Muy alta

Explique brevemente su respuesta:

F. Cierre

34. Comentarios finales, recomendaciones o ejemplos adicionales que considere relevantes para el diagnóstico del proceso actual de pricing de CDTs.

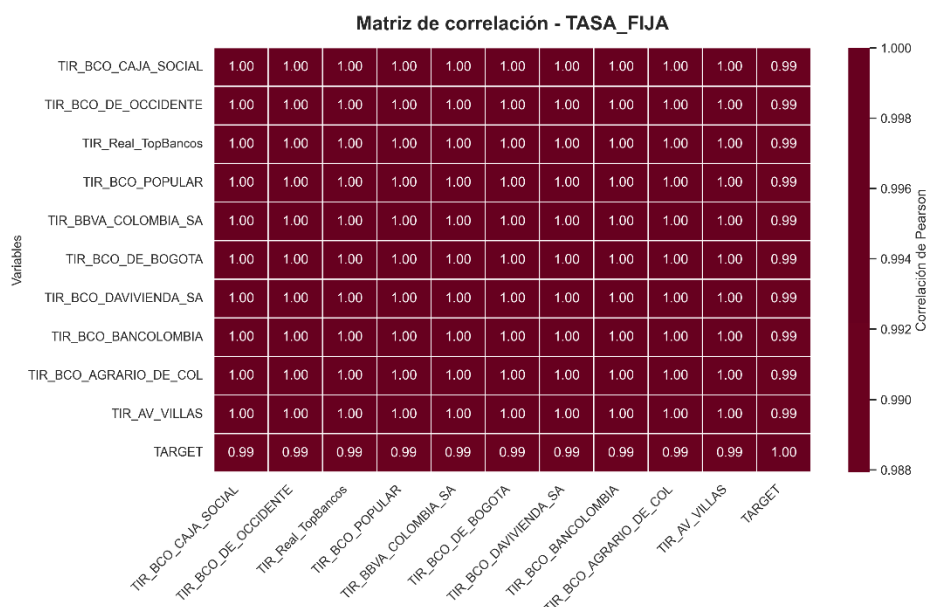
35. ¿Estaría dispuesto(a) a participar en una revisión breve de hallazgos o validación posterior del diagnóstico? ☐ Sí ☐ No

Nota metodológica. El instrumento fue diseñado para privilegiar hechos recientes y evidencia de comportamiento sobre opiniones generales, inspirado en los principios de The Mom Test (Fitzpatrick, 2013)

Anexo C - Matriz de correlación de cada escenario.

Figura 29

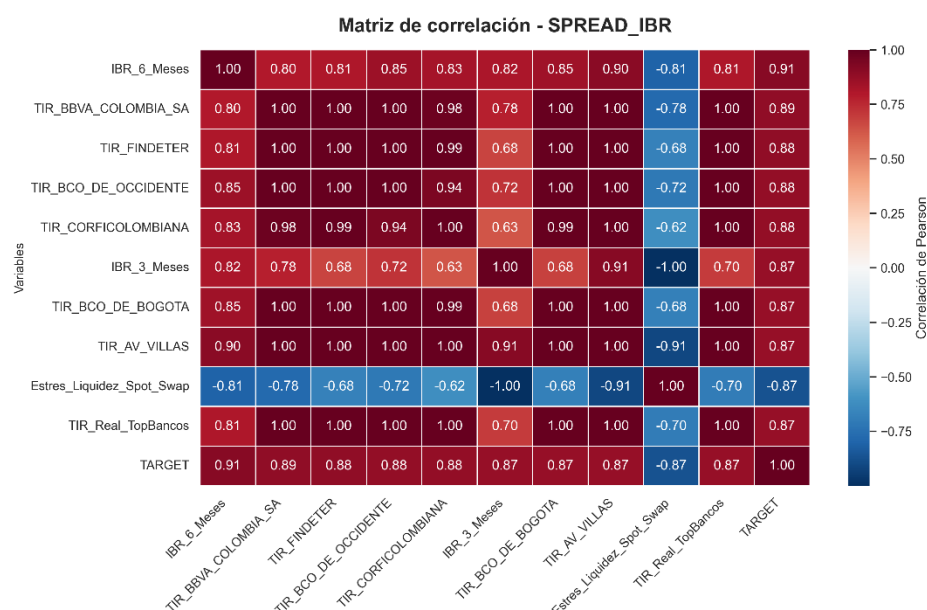
Matriz de Correlación Para Tasa Fija



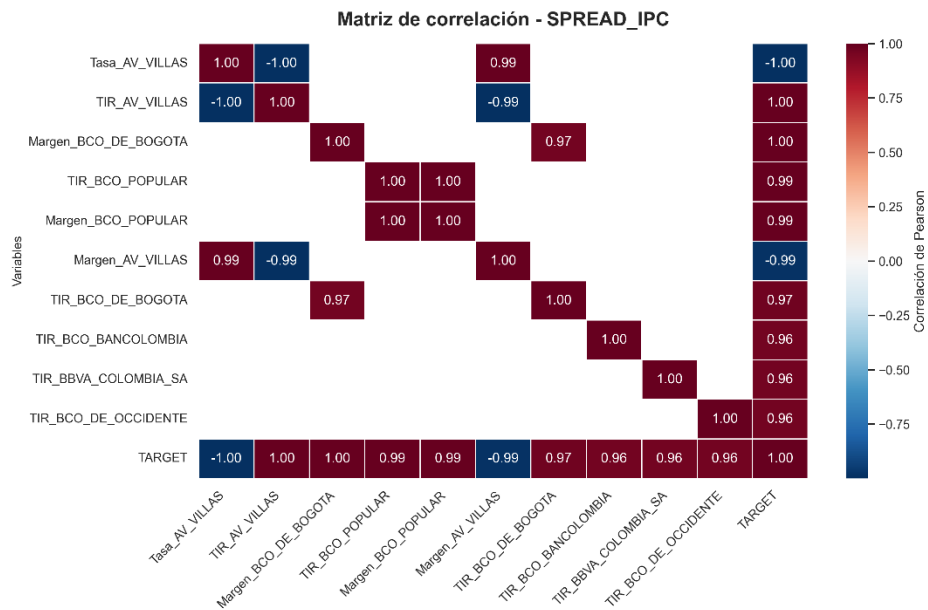
Nota. Elaboración propia a partir de la ejecución del script.

Figura 30

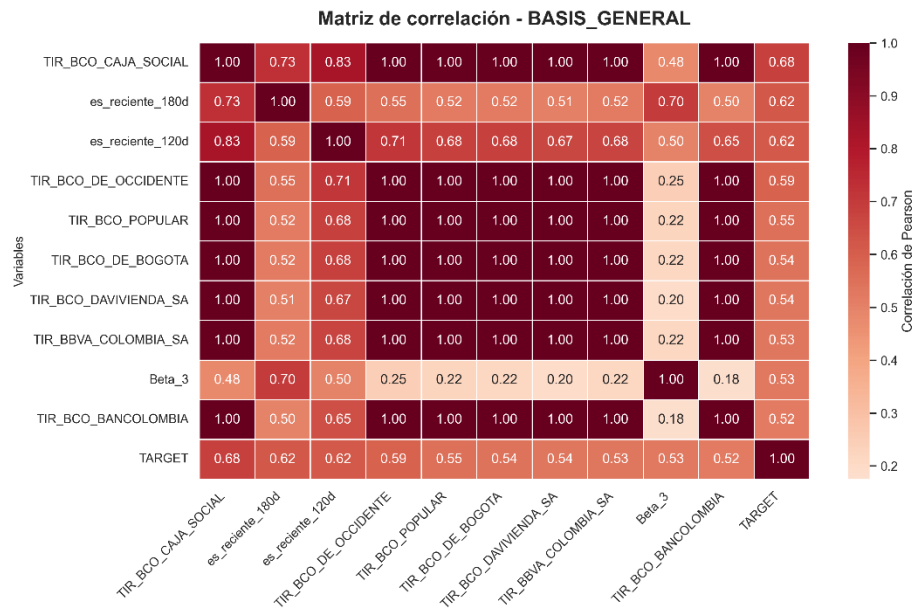
Matriz de Correlación Para IBR



Nota. Elaboración propia a partir de la ejecución del script.

Figura 31**Matriz de Correlación Para IPC**

Nota. Elaboración propia a partir de la ejecución del script.

Figura 32**Matriz de Correlación Para Basis**

Nota. Elaboración propia a partir de la ejecución del script.

Anexo D - Glosario técnico y diccionario de variables del pipeline

Tabla 34

Glosario Técnico y Diccionario de Variables del Pipeline

Término / variable	Definición operativa en el proyecto
ALM	Gestión de activos y pasivos. En el proyecto se relaciona con la administración del costo de fondeo, la liquidez, los plazos y la sensibilidad del balance frente a tasas de interés.
API SDMX Banco de la República	Servicio usado para extraer series macroeconómicas y monetarias oficiales, como IBR, TRM, UVR y tasa de política monetaria.
Banco de la República	Fuente oficial de variables macroeconómicas y monetarias utilizadas como contexto para el pricing de CDTs.
Basis de emisión	Medida relativa de pricing que compara la condición de emisión de un CDT frente a una referencia comparable de mercado. En el pipeline se calcula inicialmente con información de Precia y luego se enriquece con curvas y competencia.
Beta 0	Parámetro de curva asociado al nivel general de tasas dentro de modelos paramétricos como Nelson-Siegel o Svensson.
Beta 1	Parámetro asociado principalmente a la pendiente de la curva de tasas.
Beta 2	Parámetro asociado principalmente a la curvatura de la curva de tasas.
Beta 3	Parámetro adicional usado en modelos extendidos como Svensson para capturar formas de curva más flexibles.
BRC1+	Calificación de corto plazo utilizada como filtro de calidad crediticia en los archivos SX.
Calificación crediticia	Indicador de calidad crediticia del instrumento o emisor. En el pipeline se usa para filtrar instrumentos comparables y reducir ruido por riesgo de crédito.
CDT	Certificado de Depósito a Término. Instrumento de captación mediante el cual una entidad recibe recursos por un plazo determinado y paga una tasa o spread al vencimiento.
Convexidad	Medida de sensibilidad no lineal del precio de un instrumento de renta fija frente a cambios en la tasa de interés. Complementa la duración.
CRISP-DM	Metodología usada para estructurar el proyecto analítico en fases de comprensión del negocio, comprensión de datos, preparación, modelado, evaluación y despliegue.
Curva cero cupón	Curva que representa tasas spot para distintos plazos, sin considerar cupones intermedios. En el proyecto se usa como referencia de mercado por plazo.

Curva soberana	Curva construida a partir de TES. Se utiliza como referencia para comparar emisiones de tasa fija e instrumentos asociados a deuda pública.
Curva swap	Curva que refleja el costo de intercambiar flujos de tasa fija por flujos de tasa variable. En el proyecto se usa como referencia para CDTs indexados al IBR.
Curva_Soberana_Base	Variable construida en el feature engineering para asignar a cada CDT la referencia soberana aplicable según fecha, plazo e indexador.
Data lake	Estructura de almacenamiento donde los archivos procesados se conservan en formato analítico, principalmente Parquet.
Dias_Rezago_Precia	Diferencia entre la fecha en que Precia registra o valora por primera vez el CDT y la fecha oficial de emisión. Se usa para validar que la observación corresponda a una emisión primaria cercana a su origen.
Duración	Medida de sensibilidad del precio de un instrumento de renta fija frente a cambios en la tasa de interés.
Duración modificada	Medida ajustada de sensibilidad porcentual del precio ante cambios en la tasa.
Emisión primaria	Primera observación válida de un CDT cercana a su fecha de emisión. En el pipeline se aproxima tomando la primera aparición por ISIN con rezago controlado.
Emisor	Entidad financiera que emite el CDT o título. En el pipeline se identifica mediante campos del archivo SX y el maestro NEMOS-ISIN.
Estres_Liquidez_Spot_Swap	Variable que mide la diferencia entre la curva swap y el IBR a tres meses, usada como señal de presión relativa entre mercado monetario y derivados.
Fecha de archivo	Fecha del archivo de Precia en el que aparece el instrumento. Se usa como proxy de fecha de valoración u observación en el mercado.
Fecha de emisión	Fecha contractual en la que nace el CDT o instrumento financiero.
Fecha de vencimiento	Fecha contractual en la que vence el instrumento y se paga el capital o rendimiento correspondiente.
Feature engineering	Proceso de construcción de variables explicativas a partir de datos crudos, curvas, spreads, señales competitivas y variables macroeconómicas.
Filtro de calificación	Regla de depuración que conserva instrumentos con calificaciones de alta calidad, como AAA, F1+ y BRC1+, para mejorar comparabilidad crediticia.
F1+	Calificación de corto plazo considerada de alta calidad dentro del filtro inicial del pipeline.
FTP	Precio de transferencia de fondos. Referencia interna que permite evaluar si una captación es sostenible para el balance.
IBR	Indicador Bancario de Referencia. Tasa de referencia del mercado monetario colombiano, usada como indexador en instrumentos de tasa variable.

Indexador	Variable de referencia sobre la cual se expresa la tasa de un CDT. En el proyecto se consideran principalmente tasa fija, IBR, IPC y UVR.
IPC	Índice de Precios al Consumidor. Referencia de inflación utilizada en instrumentos indexados o reales.
ISIN	Identificador internacional del título. En el pipeline se usa para deduplicar instrumentos e identificar la primera observación de cada emisión.
Maestro CDT	Base estructurada construida a partir de archivos SX, con información de CDTs, emisor, tasa, TIR, margen, fechas, calificación e identificadores.
Maestro TES	Base estructurada construida a partir de archivos SX, con información de TES, TIR, precio, duración, vencimiento y plazo.
Margen	Spread pactado u observado sobre un indexador. En instrumentos IBR o IPC, representa el diferencial frente a la tasa de referencia.
NEMO / Nemotécnico	Código corto que identifica instrumentos en los archivos de mercado. En el pipeline se usa para clasificar TES, CDTs y emisores.
Parquet	Formato columnar utilizado para almacenar datos procesados de manera eficiente y reproducible.
PCHIP	Método de interpolación cúbica por tramos que preserva la forma de la curva y evita oscilaciones artificiales entre nodos.
Precia	Proveedor de precios y curvas de mercado utilizado como fuente principal de información de valoración, curvas y emisiones.
Premium_Pagado_Sobre_TopBancos	Variable que compara la condición de un CDT frente al promedio observado en bancos comparables.
Prima_Emission_Basis	Variable calculada inicialmente como diferencia entre TIR y tasa para tasa fija, o entre margen y tasa/spread para instrumentos indexados.
SC	Archivo de Precia que contiene curvas cero cupón por fecha, curva y plazo.
SB	Archivo de Precia que contiene parámetros o betas de curvas, usados para representar nivel, pendiente y curvatura.
Spread	Diferencial de tasa frente a una referencia. Puede calcularse frente a IBR, curva soberana, curva AAA, swap o bancos comparables.
Spread_Banco_vs_Soberano	Variable que mide la diferencia entre la TIR del CDT y la curva soberana comparable.
Spread_Vitrina_vs_TopBancos	Variable que compara la tasa facial del CDT frente a la tasa facial promedio del grupo de bancos comparables.
SwapCC	Archivo de Precia con curvas swap asociadas al IBR. Se usa como referencia para instrumentos indexados a tasa variable.
SX	Archivo de Precia que contiene información de valoración de instrumentos individuales, incluidos TES y CDTs. Es la principal fuente para construir maestros de instrumentos.

Tasa facial	Tasa pactada o registrada contractualmente para el instrumento.
Tasa fija	Modalidad de CDT en la que el rendimiento se expresa como una tasa determinada, sin depender de un indexador variable.
Tasa política monetaria / TPM	Tasa definida por el Banco de la República que sirve como referencia general para el costo del dinero en la economía.
TES	Títulos de deuda pública colombiana. En el proyecto se usan como referencia soberana para construir curvas comparables.
TIR_Real_TopBancos	Promedio de TIR observado en el grupo de bancos comparables para una fecha, tipo de tasa y plazo.
TIR	Tasa interna de retorno observada o valorada para un instrumento. En el pipeline se usa para calcular primas y spreads frente a referencias.
Top de bancos comparables	Grupo de entidades financieras usado como universo competitivo de referencia para calcular señales de mercado.
TRM	Tasa Representativa del Mercado. Variable macroeconómica usada como contexto de condiciones cambiarias.
UVR	Unidad de Valor Real. Unidad de cuenta asociada a la inflación, usada para instrumentos con componente real o indexado.
Tasa_Facial_TopBancos	Promedio de tasa facial observado en el grupo de bancos comparables para una fecha, tipo de tasa y plazo.
Tasa_Modelo_AAA_Exacta	Curva AAA aplicable según el tipo de indexador del CDT, usada como referencia de alta calidad crediticia en el dataset.
Vencimiento	Horizonte temporal hasta el pago final del instrumento. En el pipeline se usa para calcular plazo original en días.