

MODELOS DE CALIFICACIÓN CREDITICIA CON INFORMACIÓN ALTERNATIVA

Recomendaciones para su implementación en **COLOMBIA**



Modelos de Calificación Crediticia con Información Alternativa:

Recomendaciones para su implementación en Colombia

Equipo de trabajo:

Banco Interamericano de Desarrollo - (BID)

Juan Martínez Álvarez

Especialista Senior de Mercados Financieros

Luis Restrepo

Consultor

Alejandra Dussán

Consultora

Rodrigo Rocha

Consultor

Jaime Edelshein

Consultor

Departamento Nacional de Planeación - (DNP)

Mónica Lorena Ortiz Medina

Directora Técnica de Innovación y Desarrollo Empresarial

Julián Torres Jiménez

Coordinador equipo de financiamiento empresarial

Viviana Torres Ortiz

Asesora

Natalia Orozco

Asesora



Copyright © 2024 Inter-American Development Bank (“IDB”). This work is subject to a Creative Commons license CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB's name for any purpose other than for attribution and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Inter-American Development Bank
1300 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org

El Sector de Instituciones para el Desarrollo fue responsable de la producción de esta publicación.

Proveedores externos:

Edición de producción:

Editor:

Diseño: Gabriel Ramírez

Contenido

Resumen ejecutivo	ix
Antecedentes	ix
Objetivo, metodología y alcance del reporte	ix
Metodología	x
Fuentes de datos alternativas	xiii
Entorno normativo	xiii
Recomendaciones y conclusiones	xiv
1. Contexto de inclusión financiera de mipymes en Colombia	1
2. Contexto sobre las fuentes de datos alternativos	2
2.1. Datos alternativos	2
2.2. Generalidades sobre uso de datos alternativos	5
2.3. La huella digital como insumo de algunas fuentes de datos alternativos	8
2.4. Tendencias principales	9
2.5. Retos principales y criterios al evaluar fuentes de datos alternativos	10
3. Marco conceptual del diagnóstico	11
3.1. Alcance	11
3.2. Metodología de diagnóstico	14
3.3. Iniciación	14
3.4. Revisión de literatura	14
3.5. Entrevistas	15
3.6. Análisis	17
3.7. Síntesis	17
3.8. Atributos técnicos y funcionales de evaluación	17
3.9. Herramientas y técnicas para levantamiento de información	18
4. Datos de redes sociales	21
4.1. Generalidades en Colombia y en el mundo	21
4.2. Proceso de recolección o producción	22
4.3. Conversión, fricción y barreras de adopción en el proceso	24
5. Datos de psicometría	25
5.1. Generalidades en Colombia y en el mundo	25
5.2. Fundamento teórico/académico	25
5.3. Proceso de recolección o producción de datos psicométricos	28
5.4. Disponibilidad, cobertura y veracidad de datos	29
5.5. Requerimientos para la implementación	30
5.6. Resumen de beneficios potenciales	32
5.7. Resumen de retos potenciales	33
5.8. Principales proveedores	34
5.9. Análisis de propuestas comerciales	36

6. Datos de <i>open banking</i>	38
6.1. Generalidades en Colombia y en el mundo	38
6.2. Fundamento teórico	41
6.3. Modelos de negocio posibles a través del <i>open banking</i>	42
6.4. Proceso de recolección de los datos	43
6.5. Disponibilidad, cobertura y veracidad de datos	44
6.6. Resumen de beneficios y retos potenciales	45
6.7. Principales proveedores	47
6.8. Análisis de propuestas comerciales	49
6.9. Conclusiones <i>open banking</i>	51
7. Datos de telefonía móvil (Telco/CDR)	51
7.1. Generalidades	51
7.2. Experiencias internacionales	52
7.3. Resumen de beneficios potenciales	52
7.4. Resumen de retos potenciales	53
8. Datos de dispositivos <i>smartphone</i> (extraídos por SDK o similar)	54
8.1. Generalidades	54
8.2. Resumen de beneficios potenciales	56
8.3. Resumen de retos potenciales	57
9. Datos de <i>e-mail</i>	58
9.1. Generalidades	58
9.2. Resumen de beneficios en uso de datos de <i>e-mail</i>	60
9.3. Resumen de retos en uso de datos de <i>e-mail</i>	60
10. Datos de transacciones públicas	61
10.1. Generalidades	61
10.2. Experiencia que combina diferentes fuentes de datos públicos en EPC en Bogotá, Colombia. El caso Ágata y la SDDE	66
10.3. Resumen de beneficios potenciales de datos de transacciones públicas	67
10.4. Resumen de retos potenciales en datos de transacciones públicas	67
10.5. Principales proveedores	69
10.6. Análisis de propuestas comerciales	69
11. Entorno normativo	71
11.1. Entorno normativo de <i>open banking</i> en Colombia	72
11.2. Plan Nacional de Desarrollo 2023-2026	74
11.3. Normativa local e internacional de manejo de datos personales	74

12. Recomendaciones de selección y uso de fuentes de datos para el mercado EPC colombiano	74
12.1. Tácticas para mejorar la política de crédito o poder predictivo de modelos de crédito	75
12.2. Capacidades técnicas de la organización usuaria de datos alternativos	76
12.3. Definición del nivel de procesamiento de datos de los resultados que entregará el proveedor de datos alternativos	76
12.4. Combinación de fuentes de datos	78
12.5. Esquema de trabajo propuesto para el desarrollo de modelos con datos alternativos	79
12.6. Esquema de trabajo propuesto para el uso de datos alternativos considerando modelos de <i>open banking</i> existentes	80
12.7. Aplicar principios de gestión de proyectos y de sus restricciones al realizar pilotos o pruebas de concepto	82
12.8. Recomendaciones sobre la disponibilidad de los datos	86
12.9. Recomendaciones sobre sesgos y prácticas de discriminación	87
12.10. Establecimiento de criterios de éxito en la POC o piloto	87
12.11. Recomendaciones sobre el monitoreo del desempeño de los modelos alternativos	88
12.12. Incluir documentación técnica del modelo de crédito alternativo en alcance y plan de analítica	89
12.13. Sobre el uso de datos alternativos en EPC en zonas rurales	89
13. Recomendaciones sobre capacidades y acciones de la banca pública de desarrollo	90
13.1. Habilitar el uso de metodologías de modelamiento usando datos alternativos	91
13.2. Vinculación con agencias/firmas de analítica de datos y sus capacidades técnicas	91
13.3. Aproximaciones regulatorias en las primeras implementaciones de modelos con datos alternativos	92
13.4. Promoción con facilitadores de innovación como el Sandbox regulatorio	93
13.5. Adopción de estrategias de <i>marketplace</i> para la democratización de los datos para maximizar el impacto con múltiples ofertas de crédito	93
13.6. Selección y priorización de IF aliadas para implementar los esquemas de trabajo propuesto	95
13.7. Diseñar estrategias ante las principales razones de no uso o adopción de fuentes de datos alternativos	95
13.8. Educación en cultura de empoderamiento, confianza y democratización de datos	97
13.9. Asesoría técnica de apoyo a la banca pública de desarrollo	97
14. Recomendaciones de política pública	98
14.1. Segmentación de las EPC	98
14.2. Definición de producto para pilotear el uso de modelos de crédito alternativo	98
14.3. Definición de segmento o sector para pilotear el uso de modelos de crédito alternativo	99
14.4. Recomendaciones regulatorias	101
14.5. La privacidad como una serie de requerimientos normativos y un eje para el diseño de procesos de crédito y productos financieros	101
14.6. Tratamiento de los datos alternativos	102

14.7. Fortalecimiento del banco de datos	102
14.8. Habilitación de mecanismos de garantía de riesgo compartido	103
14.9. Uso de datos alternativos para desarrollar productos financieros acorde con las necesidades reales de las EPC	103
14.10. Uso de marcos que promuevan la igualdad en las oportunidades de crédito	104
14.11. Uso de marcos que promuevan el buen manejo de los datos alternativos con óptica de central de información	104
Glosario de términos	105
Bibliografía	110
Lista de figuras	
Figura 2-1. Tipos y fuentes de datos alternativos transaccionales	3
Figura 2-2. Tipos y fuentes de datos alternativos no transaccionales	4
Figura 2-3. Visión holística de los datos alternativos	5
Figura 2-4. Data Journey de datos alternativos	6
Figura 3-1. Fuentes de datos alternativos dentro del alcance de esta investigación	13
Figura 3-2. Alcance del documento por etapas de la investigación	13
Figura 3-3. Metodología del Proyecto	14
Figura 3-4. Estructura de la revisión de literatura	15
Figura 3-5. Identificación de interesados	15
Figura 3-6. Metodología para la ejecución y análisis de las entrevistas con actores clave	16
Figura 3-7. Filtrado de hallazgos	17
Figura 3-8. Cuatro fuentes de información para revisión de literatura	19
Figura 3-9. Evaluación de proveedores	20
Figura 3-10. Metodología para realizar encuestas a las instituciones financieras relevantes	20
Figura 4-1. Incorporación de redes sociales en modelo de <i>scoring</i> orientado a mipymes	23
Figura 5-1. Flujo de proceso de recolección de datos psicométricos en un score de rescate	29
Figura 6-1. Nivel de adopción de <i>open banking</i> en países con modelos liderados por el regulador	38
Figura 6-2. Estado de la regulación en países selectos de LATAM	39
Figura 6-3. Cuatro fases de la implementación de la regulación de Brasil	40
Figura 6-4. Modelo de compartición de datos basado en PSD2*	41
Figura 6-5. Modelo 1. Embedded Finance	42
Figura 6-6. Productos de terceros en <i>marketplace</i> de banco	42
Figura 6-7. <i>Banking as a service</i> , modelo	43
Figura 6-8. Flujo de autenticación simple, basado en el OBIE del RU*	44
Figura 6-9. Tres tipos de proveedores de <i>open banking</i>	47
Figura 10-1. Penetración del consumidor de las relaciones de pago de servicios públicos y telecomunicaciones	62
Figura 10-2. Flujo de datos de aportaciones	64
Figura 10-3. Tipos de datos fiscales obtenidos a través de agregadores (Belvo) de <i>open banking</i>	64
Figura 12-1. Data Journey de datos alternativos	77
Figura 12-2. Comparación de desempeño en combinación de fuentes de datos	78
Figura 12-3. Técnicas de combinación y uso de fuentes de datos. Campeón retador y matrices de puntajes	79

Figura 12-4. Esquema de trabajo para análisis predictivo con datos alternativos	79
Figura 12-5. Ilustrativo del modelo planteado, en que a través de las API de banca, se interactúa con la información financiera de usuarios bancarios mediante un agregador	81
Figura 12-6. Modelos planteados para emitir créditos basados en las fuentes priorizadas	81
Figura 12-7. Hitos de un lanzamiento de solicitud de crédito con psicometría	83
Figura 12-8. Balance disponibilidad vs. poder predictivo en las fuentes de datos	87
Figura 13-1. Aproximación evolutiva de las implicaciones de riesgo	92
Figura 13-2. Flujo de proceso en la democratización de datos y elaboración de múltiples ofertas de crédito	94

Lista de tablas

Tabla 3-1. Criterios de evaluación de fuentes alternativas para su uso en riesgo de crédito	18
Tabla 4-1. Compilación de distintos objetivos y variables usadas por red social para predecir comportamientos de mipymes	22
Tabla 5-1. Preidentificación de proveedores psicometría	35
Tabla 5-2. Escenarios de costos de instalación inicial psicometría	36
Tabla 5-3. Escenarios de costos recurrentes mensuales de psicometría	37
Tabla 5-4. Escenarios de costos recurrentes por consulta de psicometría	37
Tabla 5-5. Servicios más frecuentes entre proveedores, pagados de manera única	38
Tabla 8-2. Beneficios y retos percibidos	46
Tabla 6-3. Preidentificación de proveedores <i>open banking</i>	48
Tabla 6-4. Escenarios de costos de instalación <i>open banking</i>	50
Tabla 6-5. Escenarios de costos recurrentes mensuales de <i>open banking</i>	50
Tabla 6-6. Escenarios de costos por consulta de <i>open banking</i>	50
Tabla 10-1. Variables seleccionadas a evaluar en datos de <i>e-mail</i> *	59
Tabla 10-2. Preidentificación de proveedores de datos transaccionales	69
Tabla 10-3. Escenarios de costos de instalación de datos de transacciones públicas	70
Tabla 10-4. Escenarios de costos recurrentes mensuales de datos de transacciones públicas*	70
Tabla 10-5. Escenarios de costos por consulta de datos de transacciones públicas	70
Tabla 12-1. Consideraciones y ejemplos para tener en cuenta en el alcance del proyecto	82
Tabla 12-2. Detonadores de costo y recomendaciones para tener en cuenta en un proyecto piloto	84
Tabla 12-3. Escenarios de costos de instalación inicial en proyecto	85
Tabla 12-4. Escenarios de costos recurrentes mensuales en proyecto	87
Tabla 13-1. Posibles agencias/firmas analíticas	91
Tabla 14-1. Ejemplos de datos transaccionales y no transaccionales por tipo de industria	100

LISTA DE SIGLAS Y ABREVIATURAS

AFP	Administradora de fondos de pensiones y de cesantías
ANS	Acuerdos de niveles de servicio
API	<i>Application Programming Interface</i> (interfaz de programación de aplicaciones, por sus siglas en inglés)
Asobancaria	Asociación Bancaria y de Entidades Financieras de Colombia
ASPSP	<i>Account servicing payment service providers</i> (proveedores de servicios de pago de servicios de cuenta, por sus siglas en inglés) Administradora de fondos de pensiones y de cesantías
AUC	<i>Area under the curve</i> (área bajo una curva, por sus siglas en inglés)
Baas	<i>Bank as a Service</i> (banca como servicio por sus siglas en inglés)
Bancóldex	Banco de Desarrollo Empresarial de Colombia
BBVA	Banco Bilbao Vizcaya Argentaria S. A.
BCD	Bancos centrales de desarrollo
B2C	<i>Business to consumer</i> (empresa a consumidor, por sus siglas en inglés)
BID	Banco Interamericano de Desarrollo
BIN	Número de identificación bancaria (por sus siglas en inglés)
CCF	Cajas de compensación familiar
CDR	<i>Call Detail Record</i> (registros detallados de llamadas, por sus siglas en inglés)
CFI	Corporación Financiera Internacional (IFC, por sus siglas en inglés)
CHURN	Tasa de cancelación o abandono de un crédito
CMA	Autoridad de Mercados y Competencia del Reino Unido
CMF	Comisión para el Mercado Financiero
CNP	Cartera no productiva
Confecoop	Confederación de Cooperativas de Colombia
CONPES	Consejo Nacional de Política Económica y Social
CRA	Comisión de Regulación de Agua Potable y Saneamiento Básico
CRSP	<i>Credit Reporting Service Providers</i> (proveedores de servicio de reporte crediticio, por sus siglas en inglés)
DANE	Departamento Administrativo Nacional de Estadística
DNP	Departamento Nacional de Planeación
DIAN	Dirección de Impuestos y Aduanas Nacionales
EAAB	Empresa de Acueducto y Alcantarillado de Bogotá
ECOIA	<i>Equal Credit Opportunity Act</i> (Ley de Igualdad de Oportunidades Crediticias, por sus siglas en inglés)
EFL	<i>Entrepreneurial Finance Lab</i> (Por sus siglas en inglés), compañía que con Lenddo, formaron LenddoEFL
EPC	Economía popular y comunitaria
EPS	Entidad promotor de salud
ETB	Empresa de Telecomunicaciones de Bogotá
FCRA	<i>Fair Credit Reporting Act</i> (Ley de Reportes Crediticios Justos, por sus siglas en inglés)
FICO	Fair Isaac Corporation, por sus siglas en inglés

Finagro	Fondo para el Financiamiento del Sector Agropecuario
FTP	Protocolo de transferencia de archivos
GB	<i>Giga Byte</i>
IA	Ingreso Base de Cotización
IBC	Consultoría tecnológica y servicios IT
ICM	<i>International Conference on Material Engineering and Manufacturing</i> (Conferencia
ICMEM	Internacional sobre Ingeniería y Fabricación de Materiales, por sus siglas en inglés)
ICV	Índice de cartera vencida
ID	Identificador único de cliente
ID4D	Identificación para el Desarrollo del Banco Mundial
IF	institución(es) financiera(s)
IMF	institución(es) microfinanciera(s)
iOS	Sistema operativo móvil de código cerrado (<i>iPhone Operating System</i> , por sus siglas en inglés)
IP o IPI	Información sobre identificación personal (protocolo de Internet)
KPI	<i>Key Performance Indicators</i> (indicadores clave de desempeño, por sus siglas en inglés)
KS	Indicador de Kolmogorov-Smirnoff
KYC	Know your customer —servicio de <i>onboarding</i> — (Conozca a su cliente —inducción—, por sus siglas en inglés)
LEI	<i>Legal Entity Identifier</i> (Identificador de Entidad Legal del G20, por sus siglas en inglés)
MAU	<i>Monthly active users</i> (usuarios activos mensuales, por sus siglas en inglés)
mipymes	micro, pequeñas y medianas empresas
ML	<i>Machine Learning</i> (aprendizaje automático)
MNO	<i>Mobile Network Operator</i> (operador de red móvil, por sus siglas en inglés)
OBIE	<i>Open banking</i> (banca abierta, por sus siglas en inglés)
OFAC	<i>Office of Foreign Assets Control</i> (Oficina de Control de Activos Extranjeros, por sus siglas en inglés)
OS	Sistema operativo, por sus siglas en inglés
PAD	Programa de Acompañamiento a Deudores
P2P	<i>Peer to peer</i> (persona a persona, por sus siglas en inglés,)
PIB	Producto interno bruto
PILA	Planilla Integrada de Liquidación de Aportes
PLN	Procesamiento de lenguaje natural
PND	Plan Nacional de Desarrollo
POC	<i>Proof of Concept</i> (prueba de concepto, por sus siglas en inglés)
PPDAI	<i>Shanghai Paipaidai Financial Information Service</i> (Servicio de Información Financiera de Shanghai Paipaidai, por sus siglas en inglés)
PSD2	Directiva de Servicios de Pago (UE)
PSI	<i>Population Stability Index</i> (índice de estabilidad de la población, por sus siglas en inglés)
pymes	Pequeñas y medianas empresas
RD	<i>Research & Development</i> (Investigación y desarrollo, por sus siglas en inglés)
ROA	<i>Return On Assets</i> (Rendimiento de los activos, por sus siglas en inglés)
RS	<i>Risk Split</i> (división del riesgo, por sus siglas en inglés)

RU	Reino Unido (Inglaterra, Escocia, Gales e Irlanda)
SBS	Superintendencia de Banca, Seguros y AFP
SDDE	Secretaría Distrital de Desarrollo Económico
SDK	<i>Software development kit</i> (estuche de desarrollo de <i>software</i> , por sus siglas en inglés)
SEDPEs	Sociedades especializadas en depósitos electrónicos
Sisbén	Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales
SME	<i>Small and Medium Enterprises</i> (mipymes, por sus siglas en inglés)
SMMLV	Salario mínimo mensual legal vigente
SMS o MMS	<i>Short Message Service</i> (servicio de mensajes cortos o servicio de mensajes simples, por sus siglas en inglés)
SO	Sistema operativo
SSL	<i>Secure Sockets Layer</i> (capa de canales de seguros, por sus siglas en inglés)
Superfinanciera o SFC	Superintendencia Financiera de Colombia
S3	<i>Simple Storage Service</i> (almacenamiento de datos seguro en la nube, por sus siglas en inglés)
UE	Unión Europea
URL	<i>Uniform Resource Locators</i> (localizador uniforme de recursos, por sus siglas en inglés)
TI	Tecnología de la información
TPP	<i>3rd party provider</i> (proveedores de servicios, por sus siglas en inglés)
VaR	<i>Value at Risk</i> (valor en riesgo, por sus siglas en inglés)
VoIP	<i>Voice Over Internet Protocol</i> (voz sobre protocolo de Internet, por sus siglas en inglés)
WBG	<i>World Bank Group</i> (Grupo Banco Mundial, por sus siglas en inglés)

RESUMEN EJECUTIVO

Antecedentes

En Colombia al 2022 existen alrededor de 5,4 millones de micro, pequeñas y medianas empresas (mipymes). El Departamento Administrativo Nacional de Estadística (DANE) estima que estas empresas representan más del 90 % del sector productivo total del país, producen aproximadamente 79 % del empleo y aportan 40 % al producto interno bruto (PIB). Sin embargo, las mipymes enfrentan retos en el acceso a crédito (DANE, 2022).

La Corporación Financiera Internacional (CFI) estima que hay una brecha de financiamiento de 138 billones de pesos colombianos (USD 42 billones) y que el 45 % de las mipymes se encuentran total o parcialmente desatendidas. Esta situación, como la describe el DANE, se debe principalmente a que cuando las mipymes se acercan a solicitar un crédito este es negado, debido a la falta de historial crediticio o a un historial negativo.

Actualmente, el progreso de modelos de *score* (resultado o puntaje) de crédito basados en datos alternativos está ganando tracción tanto en economías desarrolladas como emergentes en todo el mundo. Este tipo de modelos puede ser utilizado por jugadores nuevos y por los incumbentes para evaluar a las mipymes y facilitar su acceso al crédito.

Objetivo, metodología y alcance del reporte

El Banco Interamericano de Desarrollo (BID), a través de su División de Conectividad, Mercados y Finanzas apoyó al Departamento Nacional de Planeación de Colombia (DNP) para conocer el panorama actual de los modelos de *scoring* alternativo enfocados en mipymes en Colombia y en emitir recomendaciones para implementar estos modelos.

El presente documento describe los modelos de evaluación crediticia con datos alternativos y cómo funcionan cada una de las fuentes de datos que mejor se adaptan a las mipymes, proveyendo listados de beneficios y retos de cada una. Explica también por qué puede ayudar a las instituciones financieras colombianas y a las mipymes. Además, indaga sobre la regulación vigente en Colombia que impacta directamente en la utilización de datos alternativos. Provee recomendaciones apuntadas al DNP, para construir un ecosistema de calificación crediticia alterno que sea eficaz para bancos y proveedores de datos alternativos en Colombia que pueda manejar la gestión de información, adopción y monitoreo por instituciones financieras y sugiere los pasos que se deben tomar para lanzar un piloto.

La intención de esta investigación es presentar un marco conceptual, así como insumos de diagnóstico sobre las diferentes metodologías de calificación crediticia para mipymes, usando información no tradicional. El enfoque es alrededor de siete fuentes de datos:

1. Datos de telefonía móvil (Telco/CDR)
2. Datos de redes sociales
3. Datos de dispositivos móviles (*smartphone* o teléfono inteligente) extraídos por SDK
4. Datos de psicometría
5. Datos de *open banking* (o banca abierta)
6. Datos de e-mail (correo)
7. Datos de transacciones públicas.

Como resultado de la investigación inicial, se identificaron y aplicaron criterios de calidad de datos (con enfoque en mipymes), facilidad de implementación y cumplimiento normativo para seleccionar aquellas fuentes de datos que representan una solución viable a implementarse y que resuelven las necesidades de las mipymes: datos de psicometría, datos de *open banking*, datos de transacciones públicas. Para estas fuentes se desarrolló una segunda etapa de investigación en la que se identificaron proveedores principales en Colombia, identificación de costos de implementación, estudio normativo y riesgos.

Metodología

Para la construcción del reporte se revisaron y analizaron diversas fuentes de literatura, identificación y entrevistas a actores principales en el ecosistema como representantes de la banca de desarrollo, proveedores principales, reguladores e informantes claves. Adicional, se consolidaron los hallazgos y estructuraron las recomendaciones.

Fuentes de datos alternativas

Parte 1. Reporte en el cual se describe qué son y cuáles son los beneficios y los retos de las fuentes de datos alternativos.

Las mipymes enfrentan dificultades para obtener préstamos de bancos que utilizan un enfoque convencional para evaluar su solvencia. Estas dificultades incluyen la falta de datos financieros y operativos, brechas en la regulación, y capacidades ineficaces de gestión de riesgos por parte de distintas instituciones. Por otro lado, los bancos también consideran que el enfoque convencional es ineficiente y costoso para el procesamiento de préstamos para mipymes, que involucran créditos de bajo monto. Para abordar estos problemas, se pueden utilizar varios tipos de información de proveedores de datos de terceros, conocidos como “datos alternativos”. Mediante el uso de datos alternativos, es posible determinar la solvencia, voluntad y calidad crediticia de las unidades productivas.

El objetivo general de los datos alternativos es proporcionar información adicional hacia una visión holística de una unidad productiva para la decisión del otorgamiento de un crédito. Existen diversas maneras de implementarse y los casos de uso dependen fundamentalmente de los objetivos particulares que la entidad financiera defina (por ejemplo, aumentar niveles de aprobación o mejorar el poder predictivo de modelos de crédito actuales o nuevos servicios a prestar).

Los conjuntos de datos en el mercado están en constante cambio, por lo que es difícil analizarlos y encontrar los que verdaderamente sean de utilidad. Muchos de estos datos no pueden ser evaluados, sino hasta contar con datos de prueba proporcionados por el proveedor. En este documento se definió la calidad de los datos mediante cuatro criterios: disponibilidad, cobertura, especificidad y ortogonalidad¹. Para cada una de las fuentes, se evaluaron estos criterios y con base en ese contexto, se extendió la investigación a las fuentes antes mencionadas.

	Área de enfoque	Fortalezas	Debilidades
Datos de redes sociales	Las redes sociales son populares entre las pymes como herramientas para presentarse al público y adquirir nuevos clientes: plataformas como LinkedIn y Facebook recopilan perfiles de las unidades productivas y reportan información sobre sus negocios, sus iniciativas más recientes, así como sus interacciones con sus clientes. Estos datos se usan para elaborar modelos que predicen el valor de una empresa.	<ul style="list-style-type: none"> • Alta disponibilidad. • Volumen alto de información. • Varios casos de éxito en créditos de consumo. • Herramienta para identificar voluntad de pago (inferencia de comportamiento financiero). 	<ul style="list-style-type: none"> • Dificil acceso. • Poco uso en casos de las unidades productivas, por lo que no existen recopilaciones de mejores prácticas ni estándares. • Poca precisión para medir solvencia del cliente o capacidad de pago. • Uso fragmentado de redes sociales. Se requiere un modelo muy complejo con distintos proveedores.
Datos de psicometría	En el contexto de la evaluación crediticia, las pruebas o evaluaciones psicométricas se pueden utilizar para comprender la voluntad y la habilidad de pago de un individuo. El enfoque en la voluntad de pago determina si el solicitante tiene la propensión a honrar sus deudas.	<ul style="list-style-type: none"> • Casos de éxito internacionales en préstamos sin historial. • Se demostró que en casos con deficiencia en información tradicional es viable usar esta fuente. • Se puede aumentar el uso de préstamos para pymes en el corto plazo. 	<ul style="list-style-type: none"> • Susceptibilidad de que la prueba sea manipulada. • Tiempo y completitud pueden ser una fricción para FI. • Consistencia de respuestas que dificulta evaluación.

¹ Para más información referirse al punto “2.5. Retos principales y criterios al evaluar fuentes de datos alternativos”.

	Área de enfoque	Fortalezas	Debilidades
Open banking	<p><i>Open banking</i> que permite a las FI (las instituciones financieras, financial institutions, por sus siglas en inglés) tradicionales e innovadores acceder a los datos transaccionales bancarios y no bancarios de los clientes mediante las API (por sus siglas en inglés. Interfaz de programación de aplicaciones). En todo el mundo, esta tendencia está evolucionando con rapidez. Existen distintos modelos en los que las IF pueden desarrollar productos nuevos basados en datos que normalmente no tienen acceso, pero que las API ahora permiten.</p>	<ul style="list-style-type: none"> • Industria creciente con nuevos agregadores de datos con buen número de acuerdos y penetración creciente en Colombia. • Regulación colombiana está impulsando la implementación de <i>open banking</i>. • La especificidad de los datos es alta. • Reduce costos de análisis. • Mejora el recorrido del cliente. 	<ul style="list-style-type: none"> • Cobertura limitada, debido a que el acceso a los datos depende de acuerdos que se hagan entre proveedores de datos y agregadores. • Población en Colombia altamente desatendida y no transacciona en el sistema formal, por lo que es difícil estimar la cobertura real.
Datos de telefonía móvil	<ul style="list-style-type: none"> • A través del registro del número telefónico del cliente, es posible conocer y evaluar el comportamiento que el solicitante ha tenido con su línea telefónica. • Se puede acceder a datos como: prepago/pospago, uso de datos, geolocalización, historial de recarga, patrones de llamadas y SMS, entre otros). 	<ul style="list-style-type: none"> • Análisis de riesgo asertivo. • Diseña estrategias de cobranza. • Herramienta potente para la detección de fraude. 	<ul style="list-style-type: none"> • Disponibilidad limitada, debido a regulaciones locales de protección de datos y privacidad. • Disponibilidad en declive, debido a servicios a través de internet que no se pueden rastrear fácilmente (p. ej. WA). • Dificultades al establecer convenios con telefónicas (cobertura limitada). • Especificidad limitada en el caso de mipymes.
Datos de Dispositivos Smartphone	<ul style="list-style-type: none"> • Se refiere a los datos rastreables de un dispositivo móvil, a los que se accede mediante una App instalada con el permiso del usuario (ubicación, ubicación aproximada, información personal, usos y patrones de mensajería, biblioteca multimedia, calendario y contactos, actividad en otras apps, historial de búsquedas, etc.) 	<ul style="list-style-type: none"> • Mercado colombiano está adoptando aplicaciones bancarias cada vez más, implicando disponibilidad amplia. • Tiempo de respuesta inmediata. • Adopción alta de dispositivos smartphone. 	<ul style="list-style-type: none"> • Restricciones en permisos por parte de fabricantes. • Memoria de dispositivos. La población objetivo es propensa a eliminar Apps por falta de memoria en sus dispositivos. • Confianza de usuarios. • Esfuerzos altos de implementación. • Especificidad, se tiene que evaluar al dueño del negocio. • Requerimiento de red para transmisión de datos.

	Área de enfoque	Fortalezas	Debilidades
Datos de e-mail (correo)	El uso de datos de correo se ha extendido, debido a que ingresar un correo electrónico es parte de los datos mínimos requeridos a los solicitantes de crédito. Este mecanismo evalúa la consistencia de los datos del usuario.	<ul style="list-style-type: none"> • Detección de fraude con asertividad. • Cobertura amplia, debido a creciente uso del correo electrónico. 	<ul style="list-style-type: none"> • Dificultades en estandarización de variables y resultados de las evaluaciones, ya que la actividad de correo no siempre está ligada a la actividad empresarial. • Depende de un volumen alto de datos, lo que incrementa los esfuerzos tecnológicos. • No hay casos concretos de uso en mipymes. • Foco en fraude, no en evaluación crediticia.
Datos de transacción públicas	Se refieren a los registros de las actividades comerciales entre una empresa proveedora y sus clientes. Por lo general, incluyen información relacionada con los pagos (flujo de efectivo) e información no relacionada con dinero. Las tendencias de comportamiento emanadas del análisis de esta información pueden utilizar los ingresos y el estado financiero más reciente de una mipyme. Los datos analizados son: servicios públicos, aportes a salud, pensión y riesgos, e información fiscal.	<ul style="list-style-type: none"> • En Colombia, se puede tener una cobertura y disponibilidad amplia. • Existen agregadores de estos datos robustos y con buena presencia en el país. • Puede contener un gran volumen de datos financieros y no financieros, para producir modelos y productos nuevos. • Especificidad alta, incluso en zonas rurales. • Inversión baja. 	<ul style="list-style-type: none"> • Población altamente informal y con uso alto de efectivo. • Es frecuente encontrar inconsistencias entre titulares de una cuenta y quién hace los pagos. • Acceso a los datos todavía limitado por regulación (obtención de permisos difícil y <i>Opt-in</i>, correo electrónico de suscripción).

Entorno normativo

Parte 2. Se analiza la regulación vigente en Colombia en torno al uso de datos para evaluación crediticia de las unidades productivas del país.

Aunque una de las razones por las cuales las IF vigiladas por la Superintendencia Financiera de Colombia, no han implementado este tipo de modelos de manera generalizada se relaciona con la regulación, la cual evita que una entidad incurra en pérdidas y se disminuya el valor de sus activos como consecuencia de que un deudor o contraparte incumpla con sus obligaciones. Esto obliga a manejar altos estándares que dificulten a las entidades utilizar estos nuevos modelos, los cuales conllevan a un nivel de riesgo alto.

Sin embargo, en Colombia se han dado pasos agigantados para flexibilizar la regulación y habilitar el uso de datos alternativos hacia la evaluación crediticia y producir nuevos productos y servicios. En concreto, el Decreto 1297 de julio 2022, tiene por objetivo promover la competencia, la inclusión y la eficiencia en la prestación de servicios y productos financieros, permitiendo a las IF vigiladas perfilar mejor a sus usuarios y desarrollar estrategias y alianzas con entidades de otros sectores. Así, se definen reglas aplicables para el intercambio de datos en entidades financieras protegiendo al consumidor.

El Decreto 1297 propone un avance en la regulación del modelo de *open banking* en Colombia. Define con claridad los jugadores que podrán operar bajo estas nuevas normativas y establece las reglas de funcionamiento que deberán cumplirse cuando se instaure de manera definitiva. Si bien la compartición de datos mediante *open banking* no es obligatoria, el decreto impulsa a las IF vigiladas a abrirlos.

Los datos en Colombia están protegidos por las leyes 1266 de 2008 y 1581 de 2012, bajo el cumplimiento de las cuales es posible procesar la información que los consumidores financieros autoricen expresamente.

Adicionalmente, el Plan Nacional de Desarrollo 2022-2026 aprobado integralmente, presentará avances importantes para dar lineamiento sobre esquemas de trabajo de datos abiertos.

Recomendaciones y conclusiones

Parte 3. Análisis de los hallazgos en el contexto de la población objetivo, la regulación colombiana, los proveedores y agregadores disponibles, y las capacidades de la banca de desarrollo en Colombia. Se toma la experiencia y casos de éxito globales como punto de referencia.

A través de los capítulos 12, 13 y 14, se describen una serie de recomendaciones para que las IF, aliados financieros o agencias/firmas de analítica que tienen en sus planes ser usuarios y consumidores de datos alternativos, seleccionen y usen de la mejor manera las fuentes de datos alternativos que sirvan mejor a los propósitos de otorgar créditos al segmento de las EPC.

Las recomendaciones describen lo que se encontró, como la forma más adecuada de implementar un esquema de trabajo basado en fuentes alternativas de datos. Se incluyen apartados que van desde las políticas y procesos que deben adoptar las entidades usuarias de datos, definición de las fuentes a usar basada en el nivel de procesamiento que entregará el proveedor de datos, combinación de fuentes y el desarrollo de modelos. Asimismo, se incluyen las especificaciones para la presentación de un piloto (PoC).

Posteriormente, se hace un análisis y recomendaciones de los pasos que se podrían tomar por parte de la banca de desarrollo, así como desde la regulación, para habilitar a las entidades a que usen efectivamente estas nuevas metodologías, y tengan el impacto que se plantea como origen de este documento.

1. Contexto de inclusión financiera de mipymes en Colombia

Entre las barreras que enfrentan las mipymes para el crecimiento de una de las principales es la falta de acceso a financiamiento, que se puede agudizar durante las crisis económicas al punto de volverse no solo una barrera de crecimiento sino de supervivencia. Si bien la necesidad de financiamiento es común para todos los segmentos de negocios, la imposibilidad de acceder a crédito no suele presentarse para las grandes empresas donde en la mayoría de los casos ya se tiene una vida o expediente crediticio construido.

De acuerdo con la encuesta de micronegocios del DANE de 2020; las mipymes que requieren y solicitan crédito encuentran que la ausencia de un historial crediticio es la razón principal por las que se les niega (Banca de las Oportunidades, 2020). En ese sentido a los micronegocios que se les negó la solicitud argumentan que el 17 % se debió a no contar con historial crediticio en las centrales de información financiera, y por ello al 35,2 % se les negó la solicitud por tener un reporte negativo en dichas centrales y 31,1 % por ausencia de garantías.

Resulta interesante analizar la causal de ausencia de garantías ya que esta está principalmente asociada a préstamos de alto valor asociados a financiamiento de mediano o largo plazo, sin embargo, en el análisis de la destinación de los préstamos de la misma encuesta se evidencia que la demanda de crédito es de un 86,1 % para solucionar problemas de flujo de caja en el plazo inmediato. Esto denota que hay unas brechas entre los productos de financiamiento que realmente necesitan los micronegocios y la capacidad operativas de las IF para analizar y desembolsar los recursos. Analizar e instrumentar una garantía representa un costo para las IF y este costo de adquisición en un producto de bajo monto que puede ser no rentable para la IF y eso deja un espacio para las fuentes informales en las cuales estas fuentes (familiares, amigos, "gota a gota") tienen porcentajes de aprobación mayores a los registrados por las entidades reguladas (IF e IMF) (institución microfinanciera o instituciones microfinancieras).

Existe un concepto llamado "perdidos en la mitad" y se refiere a las pequeñas y medianas empresas (pyme), cuyo acceso es muy limitado a los servicios financieros porque son demasiado pequeñas para acceder a los servicios bancarios tradicionales o demasiado grandes para beneficiarse de los microcréditos, lo cual las lleva a estar en una incómoda zona gris (Núñez Hurtado, 2016).

Asimismo, hay un problema debido a la falta de definición y caracterización pública de las EPC, ya que no está claro si el tratamiento debe ser como persona natural con actividad empresarial o como una mipyme en proceso de crecimiento. Esa ambivalencia hace muy difuso su manejo; lo que es una realidad es que la formalidad empresarial es proporcional a la calidad de información financiera tradicional, incidiendo en la facilidad de una IF para realizar en análisis de crédito; es decir, la formalidad empresarial tendría una conexión positiva con la demanda de préstamos formales (Banca de las Oportunidades, 2020).

Las capacidades técnicas financieras de los micronegocios son un factor. Al recolectar información de micronegocios y categorizándolos desde los que no tienen ninguna clase de registro hasta los que elaboran y presentan estados financieros; se encuentra que en la medida en que las microempresas tienen un nivel de contabilidad más sofisticado se registran porcentajes de acceso al crédito formal más alto (Banca de las Oportunidades, 2020).

Sin embargo, a pesar de la evidente necesidad de fuentes de financiamiento no todas las causas de exclusión tienen que ver con fallas de mercado o de oferta de productos financieros para las EPC, también existen factores como la autoexclusión, la cual debe ser analizada con mayor profundidad y se comporta muy similar tanto en créditos para personas como en créditos para microempresarios. Tal como lo establece la encuesta de micronegocios del DANE, así como la encuesta de demanda de inclusión financiera (Banca de las Oportunidades, 2022); la mayoría de los encuestados (71,4 %), que reportó no disponer de un producto de crédito ni haber hecho una solicitud durante el último año, señaló que la principal barrera fue la de no querer tener deudas (65,4 %).

En conclusión, las EPC, al ser unidades productivas muy pequeñas, experimentan dificultades para acceder al crédito distintas a las mipymes. Entre más pequeña la EPC, se observa que hay menor formalidad empresarial, menores capacidades técnicas de llevar una contabilidad, menor asociatividad, menos información financiera formal, poco o nulo historial crediticio por lo que las IF tienen poca información para poder hacer un análisis crediticio a las EPC, de acuerdo con las políticas tradicionales que consideran de igual manera principalmente datos tradicionales.

La meta del Gobierno nacional de impulsar a la economía popular al aumentar en un 6 % los ingresos de los micronegocios que logren ser atendidos (Departamento Nacional de Planeación, 2023) requiere de estrategias innovadoras para entender mejor estos micronegocios y evaluar su riesgo, y es ahí donde los datos alternativos pueden ofrecer visibilidad a las IF y a la banca pública de desarrollo sobre el segmento EPC, en el cual con las fuentes de datos tradicionales no se podría.

2. Contexto sobre las fuentes de datos alternativos

2.1. Datos alternativos

En la evaluación crediticia, los datos tradicionales o convencionales se refieren a todos aquellos datos que no se consideran en el buró de crédito o sociodemográficos y que provienen de la institución financiera o son proporcionados por el solicitante (por ejemplo, a través de una solicitud de crédito).

Por otro lado, los datos alternativos son aquellos no considerados tradicionales, pero que resultan ser útiles a pesar de no estar relacionados directamente con el comportamiento crediticio de un consumidor (FICO, 2022).

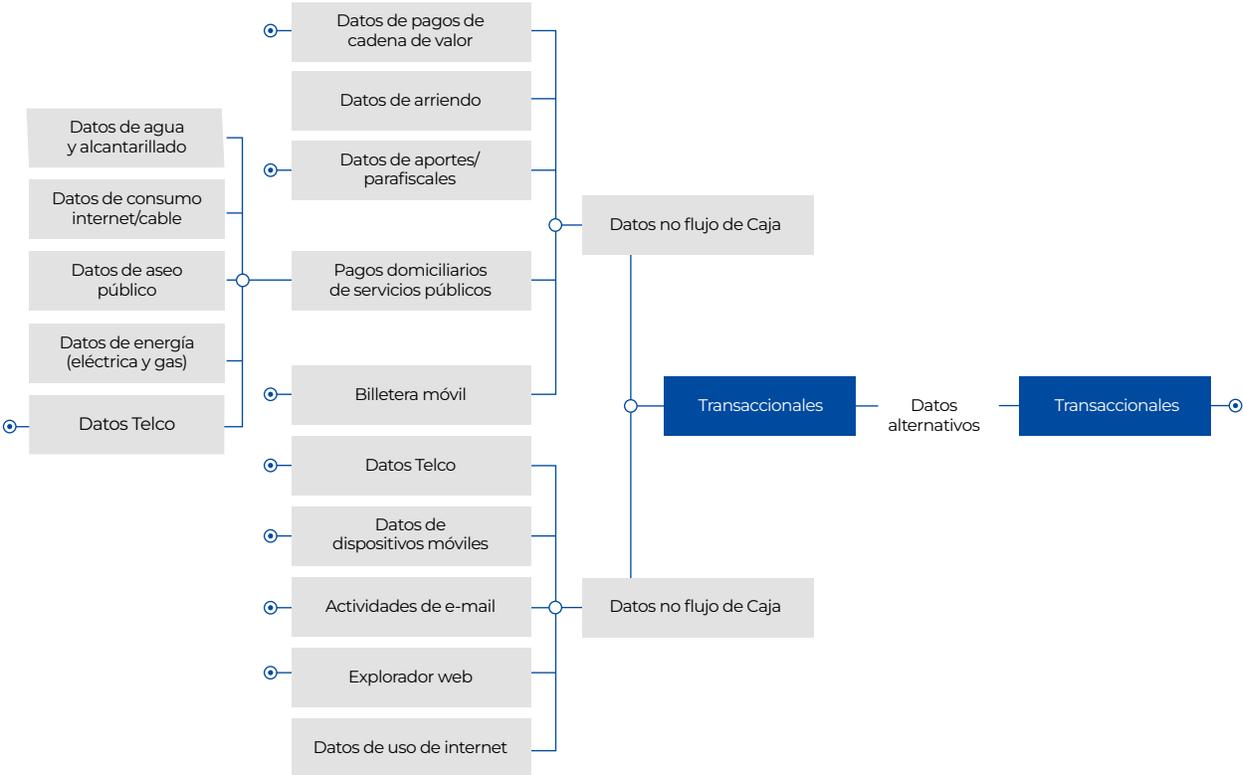
Las entidades financieras consideran que el enfoque de análisis crediticio convencional resulta ineficiente debido a la poca cobertura o información de historial crediticio, además de costoso, para el procesamiento de préstamos mipyme, que involucran montos reducidos y que no compensan el esfuerzo. Para abordar estos retos, existen varios tipos de datos no tradicionales producidos y agregados por proveedores de datos de terceros conocidos como “datos alternativos”.

Mediante el uso de datos alternativos es posible determinar la solvencia, voluntad y calidad crediticia de las mipymes. El uso de datos alternativos dentro de un modelo de evaluación crediticio es conocido como “puntuación crediticia alternativa” o como “score alternativo de crédito” (Hong Kong Applied Science and Technology Research Institute, 2020). El uso de datos alternativos como sustituto de datos tradicionales para determinar la solvencia de las mipyme es un enfoque emergente para la calificación crediticia.

El *scoring* alternativo usa datos provenientes de telecomunicaciones, empresas de servicios públicos y redes sociales; pero también puede incluir datos analizados de evaluaciones no convencionales (por ejemplo, datos basados en análisis psicométricos usados para evaluar la capacidad y voluntad de pago), o datos que rastrean actividades digitales en las redes sociales (usados para evaluar el riesgo operativo potencial de un micronegocio).

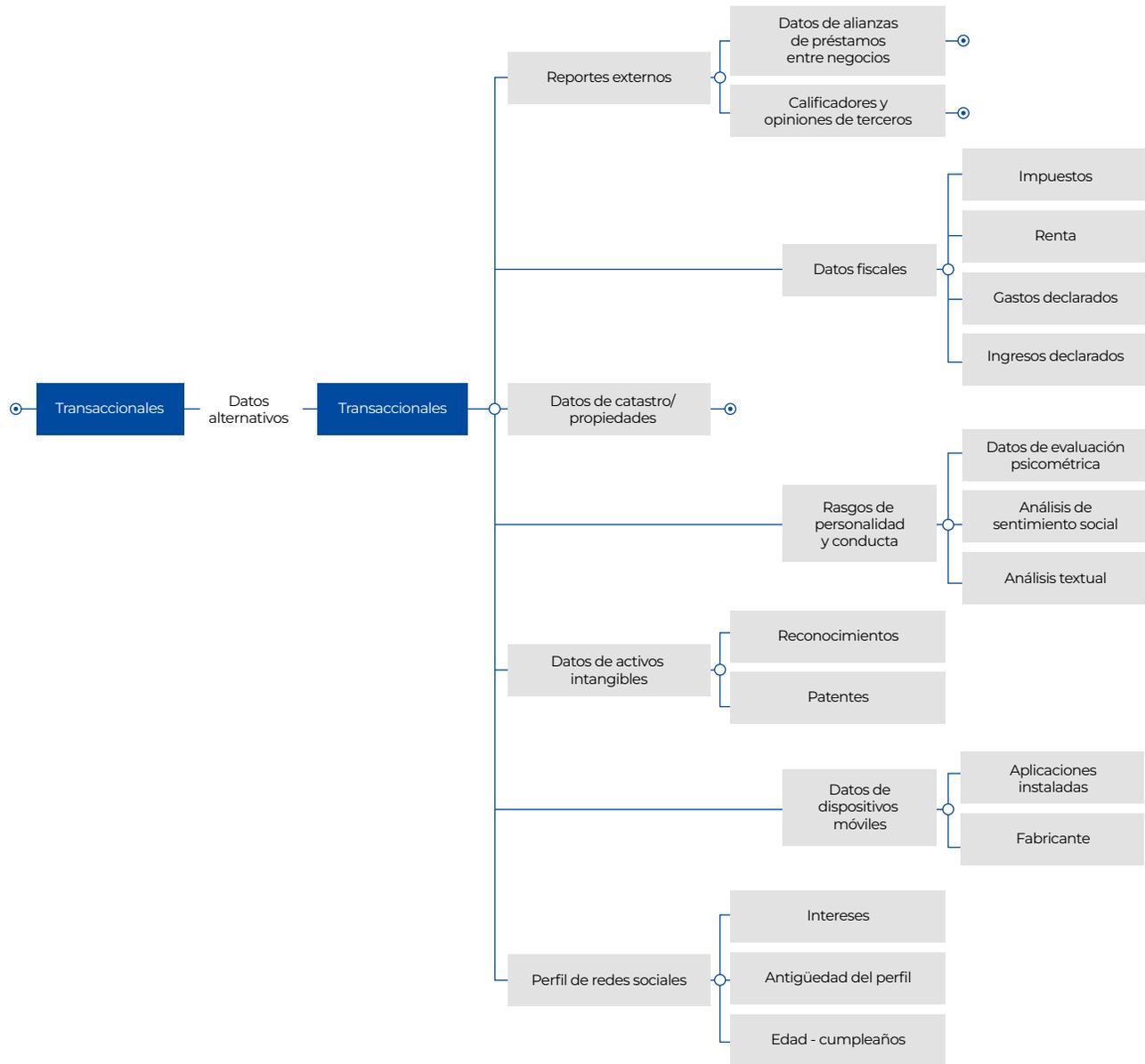
Las figuras 2-1 y 2-2 muestran algunos tipos y fuentes de datos alternativos que pueden ser usados para análisis crediticio.

Figura 2-1. Tipos y fuentes de datos alternativos transaccionales



Fuente: elaboración propia.

Figura 2-2. Tipos y fuentes de datos alternativos no transaccionales



Fuente: elaboración propia.

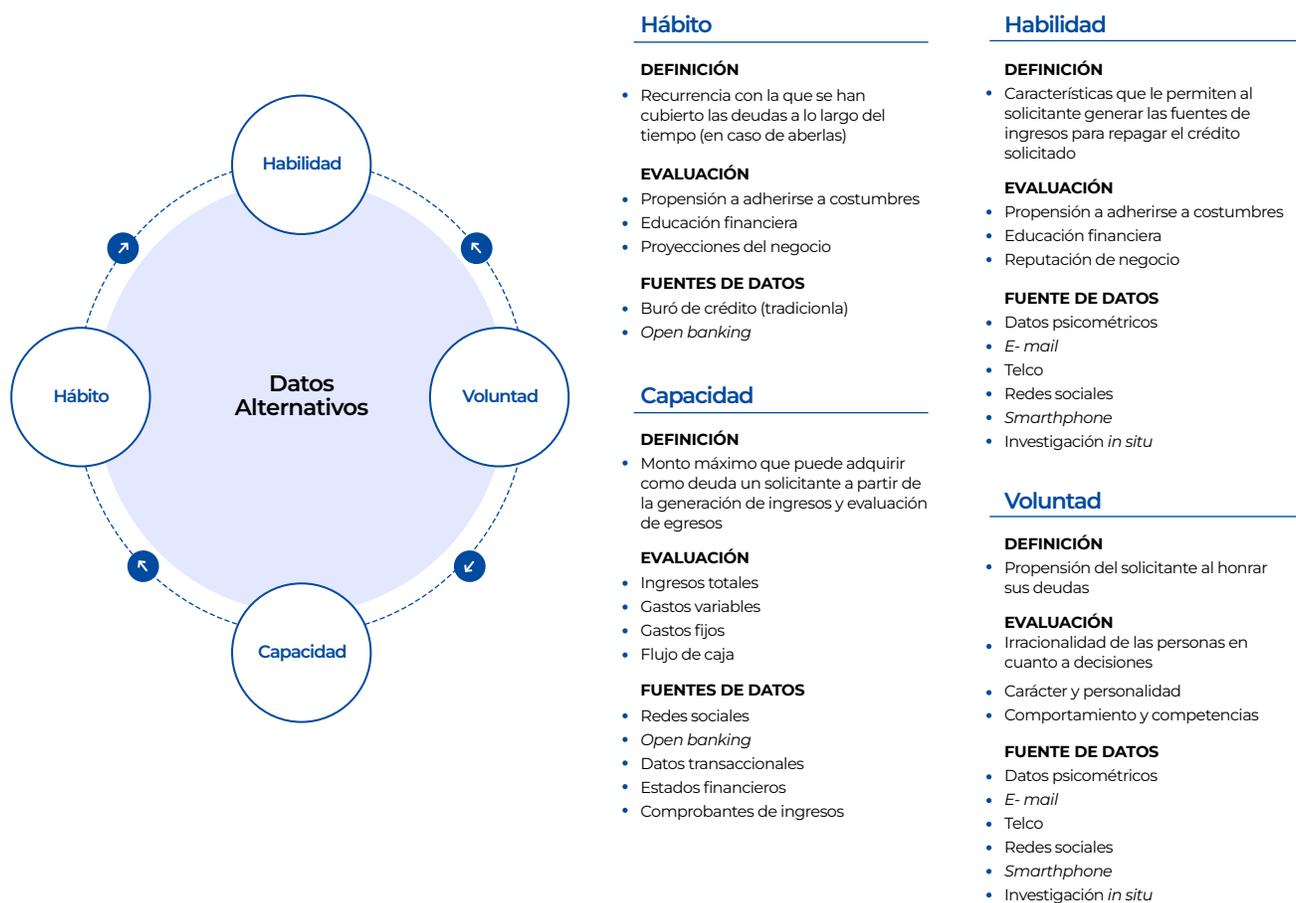
Con base en lo anterior se puede percibir que los datos tradicionales están enfocados en medir aspectos como la capacidad de pago y los hábitos de pago con evidencias transaccionales en entornos formales. A continuación, se presentan algunos ejemplos, con una óptica de datos tradicionales y una evidencia formal.

- **Capacidad de pago en una mipyme:** una declaración de renta acompañado de estados de cuentas bancarias de los últimos tres meses.
- **Hábitos de pago en una mipyme:** historia de crédito de las últimas obligaciones acompañado de un puntaje como parte de un reporte de central de información.

Como se aprecia en las figuras 2-1 y 2-2, la evaluación con datos alternativos hace aportaciones que complementan la información en cuanto a capacidad y hábito de pago en ambientes informales, brindando información sobre la voluntad de pago, así como la habilidad de este.

Derivado del tipo de datos evaluados en el *scoring* alternativo (figuras 2-1 y 2-2), es posible crear modelos enfocados a evaluar cuatro ámbitos: habilidad de pago, hábito de pago, capacidad de pago y voluntad de pago. Las fuentes de datos alternativos tienen una visión holística del solicitante del crédito en estas cuatro aristas (figura 2-3).

Figura 2-3. Visión holística de los datos alternativos

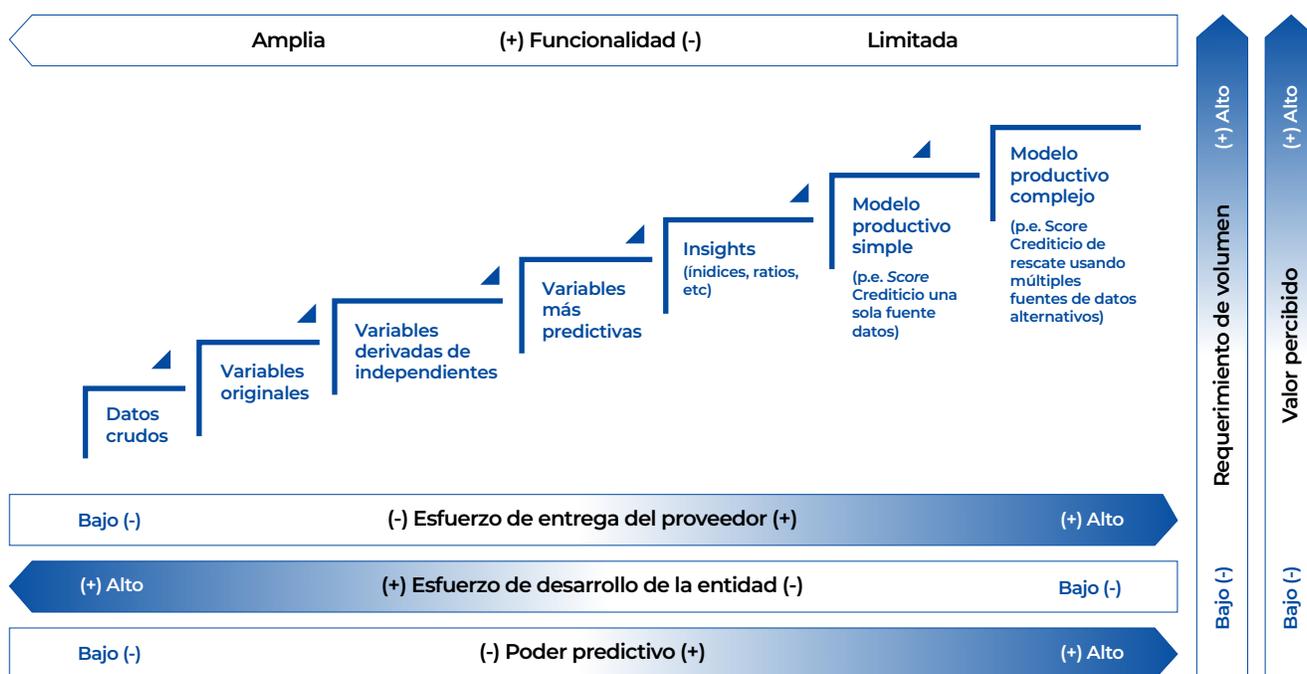


Fuente: elaboración propia.

2.2. Generalidades sobre uso de datos alternativos

Los datos alternativos pueden ser entregados de distintas maneras para derivar en diferentes tipos de valor agregado. El valor que agregan a una operación de crédito depende del grado de procesamiento que tienen y la especificidad sobre el uso para el cual serán implementados (figura 2-4). A continuación, en la figura 2-4, algunas de las modalidades de cómo son entregados los resultados de los datos alternativos.

Figura 2-4. Data Journey de datos alternativos



Fuente: elaboración propia de acuerdo con experiencia del equipo consultor.

- a. **Datos crudos:** los datos se consumen de manera directa del proveedor de la fuente sin procesamiento o tratamiento previo. Está en potestad de la entidad financiera o la agencia/firma de analítica de datos establecer los protocolos de su gobernanza, así como su procesamiento para convertir el “dato” en “información” útil para el proceso de crédito. Esto posibilita a la entidad explorar una amplia gama de casos de uso dentro del proceso; un ejemplo para datos de Telco y también para datos de dispositivos móviles, podría ser un conjunto de datos completo de llamadas telefónicas recibidas.
- b. **Datos semiprocesados:** datos después de haber tenido un procesamiento, es decir, extraídos, organizados, preanalizados y presentados para su uso posterior. Las formas más comunes de recibir estos datos se tienen en modalidades “variables” (también llamada “variable estadística”) (glosario de términos), que por lo regular son insumos para estadísticos descriptivos y posteriormente utilizados en modelos. Al continuar con las fuentes de datos de Telco y dispositivos móviles, un ejemplo podría ser la variable calculada “número promedio de llamadas diarias realizadas en los últimos tres meses”.
- c. **Insights (glosario de términos):** datos procesados que se entregan a manera de calificaciones por “atributos”. A diferencia de los datos semiprocesados, los *insights* propician una opinión o decisión respecto al solicitante del cual se obtuvo la información. Se entrega a manera de índices o rating (audiencias), que describen a ese individuo dentro de una población para una futura decisión. Al continuar con las fuentes de datos de Telco y dispositivos móviles un ejemplo podría ser “El usuario está dentro del 7 % que realiza más llamadas telefónicas”.

- d. **Scores:** puntaje resultante de calificación crediticia (o *credit scoring*), que indica la calidad crediticia del solicitante y que está asociada generalmente a una probabilidad de *default* o probabilidad de “no pago” a mayor probabilidad, menor puntaje. Esta probabilidad es resultado de un modelo estadístico predictivo que generalmente es de tres dígitos.

El objetivo general de los datos alternativos es proporcionar información adicional que permite tener una visión holística de una mipyme para la decisión de un crédito. Existen diversas maneras de implementarse y los casos de uso dependen fundamentalmente de los objetivos particulares que la entidad financiera defina. Aquí algunos ejemplos:

Estrategia 1. Aumentar niveles de aprobación: elevar la cantidad de mipymes que son aprobadas en comparación con el uso único de metodologías de calificación crediticia tradicional. En este caso se toman los solicitantes que han sido rechazados por una causal de negación “blanda” (canales de negación “blanda”, glosario de términos), es decir, los que no fueron rechazados de manera contundente por un incumplimiento flagrante a la política. Por ejemplo, el que un representante legal o microempresario aparezca en la lista OFAC (Office of Foreign Assets Control, por sus siglas en inglés, glosario de términos), es una causal contundente de negación; sin embargo, el no disponer de información de alguna central de información es considerada una “causal blanda de negación”. En este escenario, la solicitud de crédito puede ser reevaluada con la óptica de datos alternativos a través de los resultados entregados anteriormente, cuando se usa un *score* para realizar esta labor de reevaluación o reconsideración de una solicitud previamente rechazada o negada anteriormente por la política tradicional se le llama “*score de rescate*” (fuente, elaboración propia de acuerdo con experiencia del equipo consultor).

Estrategia 2. Mejorar la política o poder predictivo de modelos de crédito actuales: los datos alternativos pueden ser utilizados para tener subconjuntos de datos “adicionales” a los que ya recolecta una IF, permitiendo de esta manera enriquecer sus procesos de otorgamiento de crédito y de utilizar para elevar el poder predictivo de un puntaje de crédito tradicional. Una de las razones por las que los datos alternativos pueden ser valiosos en combinación con los tradicionales, tanto de la misma IF o los de buró, tiene que ver con la baja correlación que podría existir entre ellos, lo cual impactaría positivamente al mejorar la ortogonalidad de los datos, lo que es un atributo muy valioso en la calidad de estos, que se describirá con detalle más adelante.

Soluciones como LiftPremium de Experian² utilizan datos alternativos en combinación con tradicionales. LiftPremium afirma que puede ofrecer hasta un 10 % de aumento en comparación con los modelos tradicionales, incluido un aumento de hasta 2 % en solicitantes de “expediente grueso” (Burrows, 2023), es decir, que tienen historial crediticio para poder ser evaluados con políticas tradicionales.

² Experian es una empresa de servicios de información global con presencia en 43 países, entre ellos, Colombia. Afirma ser el buró más grande del mundo en ventas.

Estrategia 3. Mejorar la eficiencia del proceso o del canal de admisión: los datos alternativos pueden contribuir a mejorar la eficiencia de un canal de venta o de servicio de una entidad financiera. Esta eficiencia puede evaluar con diferentes ópticas como, por ejemplo, tiempo de respuesta, índice de conversión, costo de adquisición entre otras. Los datos alternativos pueden contribuir a estas métricas, en especial en canales digitales donde una campaña puede llegar a atraer muchos solicitantes, sin embargo, si no se tienen los mecanismos de evaluación adecuados se pueden tener tasas muy altas de desistimiento o negación, lo cual impacta a la eficiencia general del canal. Las IF han aumentado su eficiencia con datos alternativos al disminuir el tiempo de aprobar o negar una solicitud aumentando de esa manera su eficiencia. Casos de IMF como Te Creemos³, donde el modelo de negocio requiere de oficiales o asesores de crédito en el campo, logró aumentar la eficiencia de esta fuerza comercial en un 28 % al poder recolectar datos psicométricos *in situ*, lo que les permitió duplicar la cantidad de créditos otorgados manteniendo el mismo nivel de riesgo. Lograron mejorar el costo de adquisición al grado que pudieron mejorar su oferta económica a sus clientes del sector popular (EFL, 2017).

2.3. La huella digital como insumo de algunas fuentes de datos alternativos

La obtención de datos alternativos se encuentra principalmente en la huella digital del solicitante. La huella digital, se refiere a datos trazables que están vinculados a un individuo. Esto podría ser en línea, como historial de navegación, en una página web o interacciones en una web *app*, pero también a través de dispositivo digital como una *Device App* de un teléfono Android o iOS.

La “huella digital” también puede referirse a acciones, comunicaciones o contribuciones que pueda realizar un individuo en un entorno digital como por ejemplo “dar like”, realizar un post en una red social, enviar un correo electrónico, añadir un contacto a la libreta de direcciones, etcétera. La huella digital no es una fuente de datos alternativos independiente, sino todos aquellos datos trazables de diferentes tipos de fuentes de datos alternativos. La huella digital es cada vez más apreciada dentro de las diferentes fuentes de datos alternativos, ya que cada individuo tiene una huella digital única, sin embargo es importante considerar cambios normativos, como en el caso de las *cookies* (ficheros), que son usadas como fuente de datos para rastrear la huella digital de usuarios y que no estará disponible después de 2023 en países como México (Twilio, 2022); esto supone un gran impacto ya que alrededor del 81 % de las empresas depende en su totalidad o en gran medida de las *cookies* de terceros para enriquecer sus procesos comerciales o de servicio al cliente.

³ Te Creemos es una institución microfinanciera en México, que ofrece productos financieros para el sector popular.

2.4. Tendencias principales

Nuevos jugadores como proveedores de fuentes de datos alternativos, Fintech (glosario de términos), prestamistas alternativos y nuevos usuarios de datos (p. ej. agencias/firmas de analítica de datos, *marketplaces* (glosario de términos), e implementadores de *leads*, entre otros) se unen diariamente al ecosistema de reporte de crédito (The World Bank Group, 2022). En especial, los proveedores de datos alternativos rápidamente están surgiendo como un nuevo tipo de proveedores de servicios de reporte de crédito o centrales de información, ya que dichas organizaciones disponen en general con buenas capacidades analíticas en sus fuentes de datos no tradicionales que pueden ser usadas en casos de uso de riesgo de crédito como, por ejemplo, construcción de modelos de crédito y producción de puntajes.

En las últimas dos décadas, se han propuesto muchos modelos de calificación crediticia para las mipymes, basados en *machine learning* (ML) y un conjunto consistente de predictores, o índices financieros derivados de datos contables (por ejemplo, índice actual, rendimiento de activos, ROA, índice de deuda a capital, etcétera).

A pesar de su poder predictivo, los índices financieros tienen limitaciones. Es difícil para los prestadores de servicios financieros, obtener datos sobre el estado financiero de las mipymes. Además, puede haber otros factores más allá de los datos financieros que pueden influir en el comportamiento de incumplimiento de las mipymes, como su disposición y capacidad para pagar el préstamo (Putra, Bikash, Redi, y Bozzon, 2020).

Por esta razón, algunos estudios han comenzado a incluir predictores no financieros que incorporan conocimientos sobre comercialización, ventaja tecnológica, calidad de gestión, edad, tamaño, tipo de industria y área geográfica. Si bien mejora la precisión de la calificación crediticia, estos datos son difíciles de obtener, ya que los prestadores deben gestionar una encuesta en el sitio o tener una relación histórica lo suficientemente larga con los prestatarios para recopilar los datos (Angilella, 2015).

Otra tendencia digna a ser mencionada se refiere a que los microempresarios y los individuos en general cada vez son más conscientes sobre el valor que tienen sus datos y desean elegir cuándo y cómo estos pueden ser aprovechados para ayudarles a cumplir con sus objetivos tanto personales como comerciales. En el contexto de la inclusión financiera en las EPC, se busca que exista “democratización del crédito” y eso se puede lograr a través de la “democratización de los datos”, de manera que hay una tendencia de los microempresarios a cada vez estar más interesados en tener el poder como consumidores en cuanto a decidir sobre quiénes tendrán acceso a sus datos y como serán usados para tener productos financieros más significativos para ellos.

2.5. Retos principales y criterios al evaluar fuentes de datos alternativos

Existen desafíos en el ejercicio del descubrimiento y la evaluación de datos que incluyen una serie de barreras operativas y técnicas (Exabel, 2022).

- **Calidad de los datos**⁴: los conjuntos de datos en el mercado están en constante cambio, por lo que es difícil analizarlos y encontrar los que verdaderamente son de utilidad. Muchos de estos datos no pueden ser evaluados sino hasta tener de prueba proporcionados por el proveedor. En este documento se define la calidad de los datos mediante cuatro criterios:
 - **Disponibilidad**: los datos alternativos deben ser de fácil acceso y en la cantidad y vigencia adecuada. Es decir, se deben tener suficientes datos actualizados y que existan prácticas que los renueven con frecuencia.
 - **Cobertura**: una fuente de datos alternativa debe tener una cobertura amplia sobre el segmento o población a analizar y que sea consistente. Por ejemplo, si se refiere a datos de servicios públicos sería si la población que dispone de datos de cierto servicio público con determinada compañía, si es de datos de dispositivo móvil con Android la cobertura podría ser la penetración de este tipo de sistema operativo (OS, por sus siglas en inglés) dentro del total de la población a ser evaluada.
 - **Especificidad**: la fuente debe proporcionar información específica sobre el individuo a ser evaluado. Para el caso de EPC deberá ser específica sobre su actividad económica.
 - **Ortogonalidad**: idealmente, los datos alternativos deberían complementar los datos de los burós de crédito tradicionales. Estadísticamente es deseable que exista una baja correlación entre modelos de scores tradicionales en comparación de los alternativos. Esto se debe a que si ambos modelos tuvieran una correlación alta y similar de poder predictivo entonces el modelo alternativo sería redundante.
- **Cumplimiento normativo**: tanto compradores como vendedores potenciales deben pasar diferentes etapas de debida diligencia para entender el *Data Journey* (recorrido de los datos), desde ¿cómo es producido? hasta ¿cómo podría ser consumido? para identificar requerimientos de cumplimiento normativo. Los principales aspectos para considerar tienen que ver con privacidad y tratamiento de datos personales.
- **Contratación**: negociar contratos individuales desde cero con cada proveedor de datos requiere de tiempo. Según el tamaño de la organización contratante, así como el nivel de complejidad de las políticas de adquisición. Generalmente las organizaciones experimentan cuellos de botella en las áreas legales.

⁴ Para efectos de este documento, el término “calidad” se refiere a cuatro criterios: disponibilidad, cobertura, especificidad y ortogonalidad de los datos.

- **Ingeniería de datos:** la ingesta de datos, la limpieza, el mapeo de entidades y el modelado de datos requieren de un tiempo valioso de los recursos especializados. La eficiencia en esta actividad depende en gran medida de las capacidades técnicas de quien la realice, ya sea este el proveedor de los datos, el comprador o un intermediario como lo sería una agencia/firma de analítica de datos.
- **Herramientas analíticas:** en ausencia de herramientas dedicadas y diseñadas específicamente, el proceso de identificar el valor potencial en un conjunto de datos es en ocasiones extenuante y repetitivo, encadenado a lo largo de una serie de recursos y tecnología con poco tiempo. La eficiencia en esta actividad depende en gran medida de las capacidades técnicas de quien la realice, ya sea este el proveedor de los datos, el comprador o un intermediario como lo sería una agencia/firma de analítica de datos.
- **Riesgo operativo:** hay un costo derivado de la incertidumbre en la compra de datos y los compradores desean la mayor certeza antes de realizar la inversión. Este riesgo operativo está asociado con todos los retos y criterios mencionados anteriormente, por lo que si el proveedor o el comprador, o ambos, implementan acciones mitigantes para cada uno el riesgo residual puede ser menor.

3. Marco conceptual del diagnóstico

3.1. Alcance

La presente es una investigación que resume el diagnóstico y recomendaciones sobre las metodologías de calificación crediticia usando datos alternativos.

Al principio del ejercicio se estableció un alcance general para mipymes en Colombia y posteriormente se decidió enfocar la investigación a la economía popular comunitaria (EPC) principalmente en geografías urbanas.

Los objetivos de la presente investigación son:

- Recopilar experiencias internacionales en materia de metodologías de calificación crediticia para EPC, utilizando información no tradicional.
- Revisar experiencias sobre las evaluaciones y resultados obtenidos en términos de inclusión financiera de EPC, no bancarizadas o excluidas del acceso al financiamiento.
- Establecer requisitos para la implementación y éxito de las experiencias. Identificación de las barreras para la adopción.
- Identificar ventajas comparativas en términos de costos para las instituciones usuarias de la tecnología.

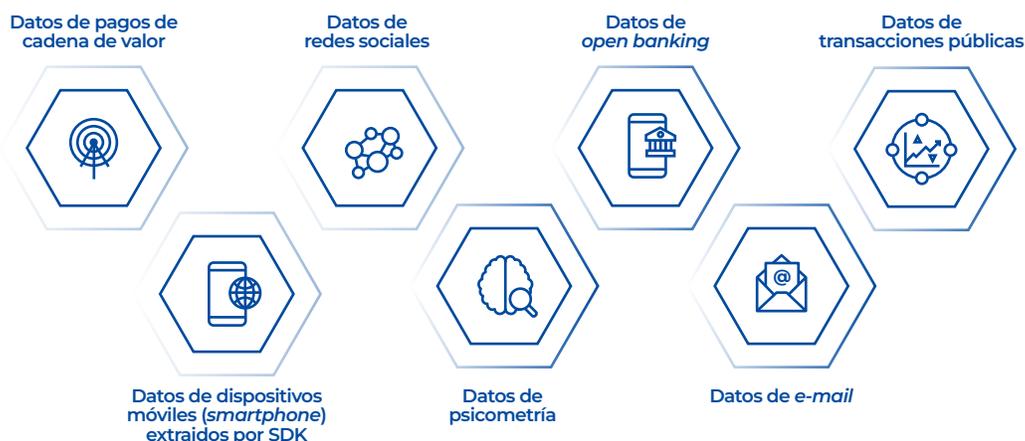
- Analizar y comparar las metodologías de calificación crediticia para EPC utilizando información no tradicional, de tal manera que se ofrezcan recomendaciones sobre cuáles serían las más apropiadas y viable para aplicar en Colombia, teniendo en cuenta el costo-beneficio de utilizar cada una de las metodologías y el acceso a la información en el país.
- Identificar requerimientos normativos y de la agenda nacional, y necesidades de acuerdos y convenios de colaboración interinstitucionales para el uso y compartición de la información si fuera necesario.
- Revisar las capacidades internas de bancos centrales de desarrollo (BCD) en Colombia, para el uso de la tecnología de calificación crediticia alternativa y la transformación digital en términos de capacidades y recursos requeridos.
- Elaborar listado de conclusiones y recomendaciones de política pública que se deben implementar desde el sector público, para el diseño y promoción de metodologías de calificación crediticia con información alternativa, acorde al contexto del país y considerando los resultados encontrados en esta investigación.

Con base en lo anterior, este documento representa un diagnóstico sobre las diferentes fuentes de datos y de proveedores de las fuentes, así como recomendaciones en tres grandes frentes de trabajo para que las IF seleccionen y usen las fuentes de datos alternativos; sobre ¿cuáles acciones? o ¿cuáles recomendaciones se deben tener en cuenta en la política pública? De manera que, cumpliendo con los objetivos anteriores, estos diferentes actores tengan en cuenta dichas recomendaciones al elaborar su plan estratégico y sus hojas de ruta en el momento de lanzar un piloto. Con base en ello, esta investigación, no se trata de un documento técnico orientado a la elaboración de modelos predictivos de crédito.

De la misma manera, el enfoque de la investigación es sobre el análisis de crédito con óptica de riesgo, en el momento de la elaboración, por lo que no considera lo referente a riesgo operacional, modelos de fraude, como tampoco recomendaciones relativas a gestión de cobros o cartera.

Las fuentes de datos alternativos que se han decidido estudiar en el presente se recogen en la figura 3-1.

Figura 3-1. Fuentes de datos alternativos dentro del alcance de esta investigación



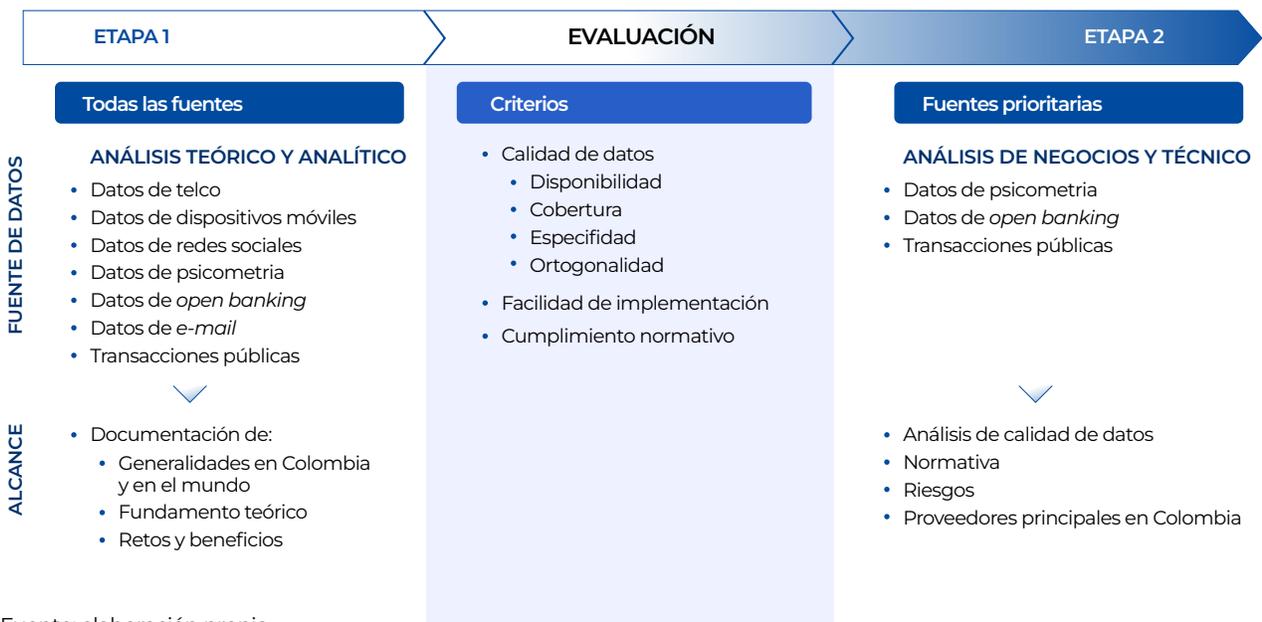
Fuente: elaboración propia.

3.1.1. Estructuración de la investigación y alcance del documento

Con el fin de presentar un reporte relevante a mipymes en Colombia y para exhibir las recomendaciones, se decidió que la investigación se dividiera en dos etapas (figura 3-2). Esto se hizo con el fin de tener un entendimiento general sobre las fuentes de datos y evaluar la relevancia de cada fuente para este proyecto, y posteriormente, realizar un análisis a profundidad solo de aquellas fuentes que fueran consideradas como prioritarias (como se muestra en la figura 3-2).

Por lo tanto, el alcance del presente documento abarca hasta el análisis teórico de todas las fuentes, de negocio y técnico para las fuentes seleccionadas. Así mismo, las recomendaciones se hacen respecto a las fuentes prioritarias.⁵

Figura 3-2. Alcance del documento por etapas de la investigación



Fuente: elaboración propia.

⁵ Los criterios de evaluación y selección de las fuentes de datos se presentan en la sección 3.8.

3.2. Metodología de diagnóstico

Para cumplir con los objetivos, se desarrolló una metodología basada en cinco etapas que se adecuaron a las actividades clave del proyecto (figura 3-3).

Figura 3-3. Metodología del Proyecto



Fuente: elaboración propia.

3.3. Iniciación

Durante la iniciación del proyecto se conformó el acta del proyecto y se sostuvieron reuniones de trabajo con todos los involucrados. Durante esta etapa, se diseñó el cronograma y la metodología.

3.4. Revisión de literatura

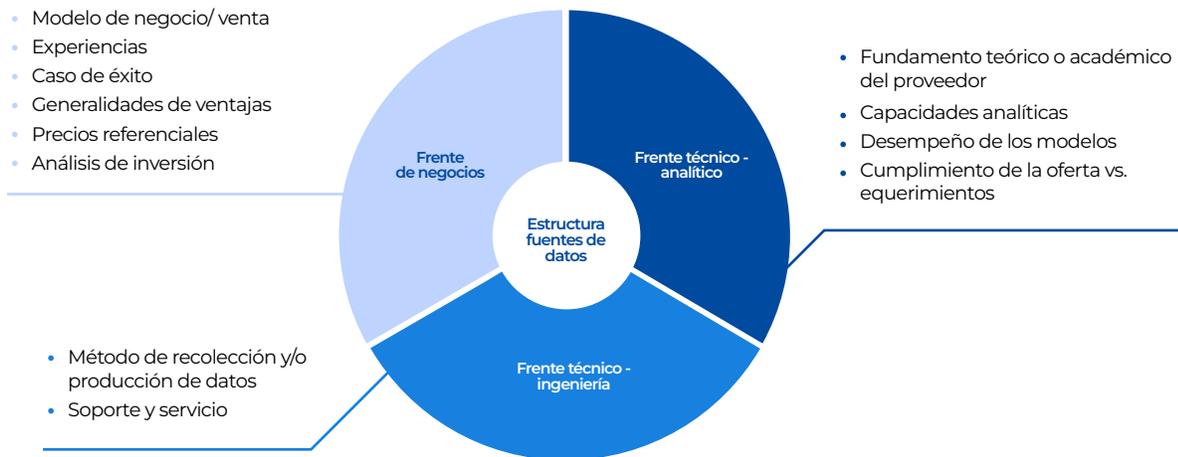
Durante esta etapa se identificaron y revisaron documentos claves para entender las estructuras de las fuentes de datos. El enfoque se centró en definir las fuentes y en determinar su relevancia para este proyecto mediante análisis bibliográfico.

El análisis de la literatura se organizó en tres frentes (figura 3-4). Se recalca que el punto A se llevó a cabo para todas las fuentes, mientras que los B y C solo para las fuentes prioritarias.

- a. **Frente teórico/Analítico**, con el fin de describir el concepto sobre el cual se construye el modelo de cada fuente de datos, así como los alcances y retos teóricos. De esta manera, se identificó el nivel de cumplimiento contra los requerimientos de este proyecto.
- b. **Frente de negocios**, con el fin de describir el trasfondo e impacto que tiene cada fuente de datos en la implementación de proyectos de *scoring* alternativo. El enfoque fue la identificación de casos de éxitos locales e internacionales, ventajas, retos y costos.

- c. **Frente técnico e ingeniería**, con el fin de describir los métodos de recolección y explotación de datos. Esto determinó el nivel de servicio de soporte necesario.

Figura 3-4. Estructura de la revisión de literatura



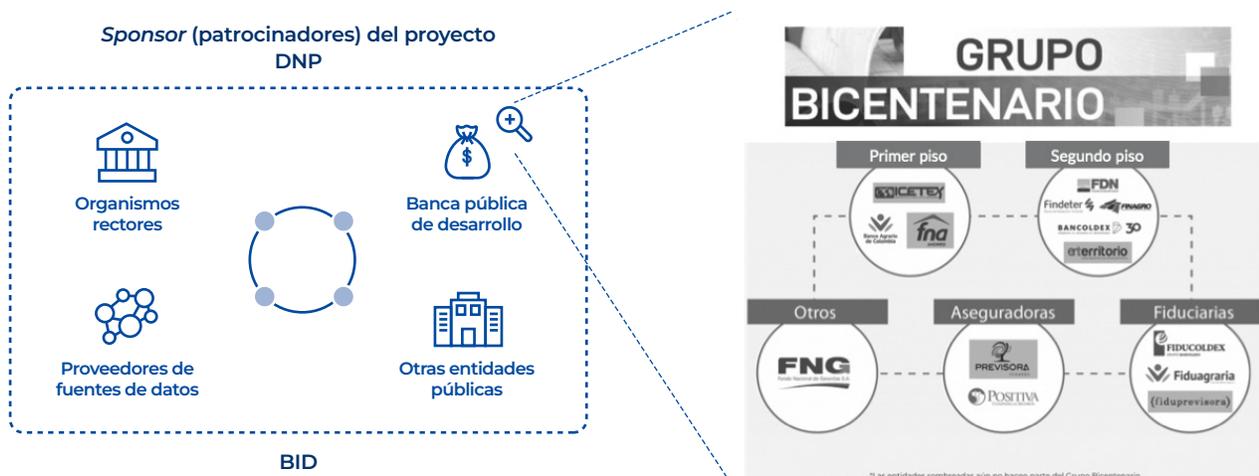
Fuente: elaboración propia.

3.5. Entrevistas

Se condujeron una serie de entrevistas con actores clave que enriquecieron los hallazgos del estudio teórico. Además, de estas entrevistas se obtuvieron distintos tipos de recomendaciones: (a) Recomendaciones de proveedores de fuentes de datos y (b) Recomendaciones de política pública.

Con esto en mente, para conducir las entrevistas y recabar datos, se identificaron interesados claves (figura 3-5), en cuatro cohortes recomendados por los patrocinadores del proyecto (DNP y BID). Esto permitió tener visibilidad de los intereses y elaborar estrategias para recabar datos y gestión de expectativas.

Figura 3-5. Identificación de interesados



Fuente: elaboración propia y página web Grupo Bicentenario.

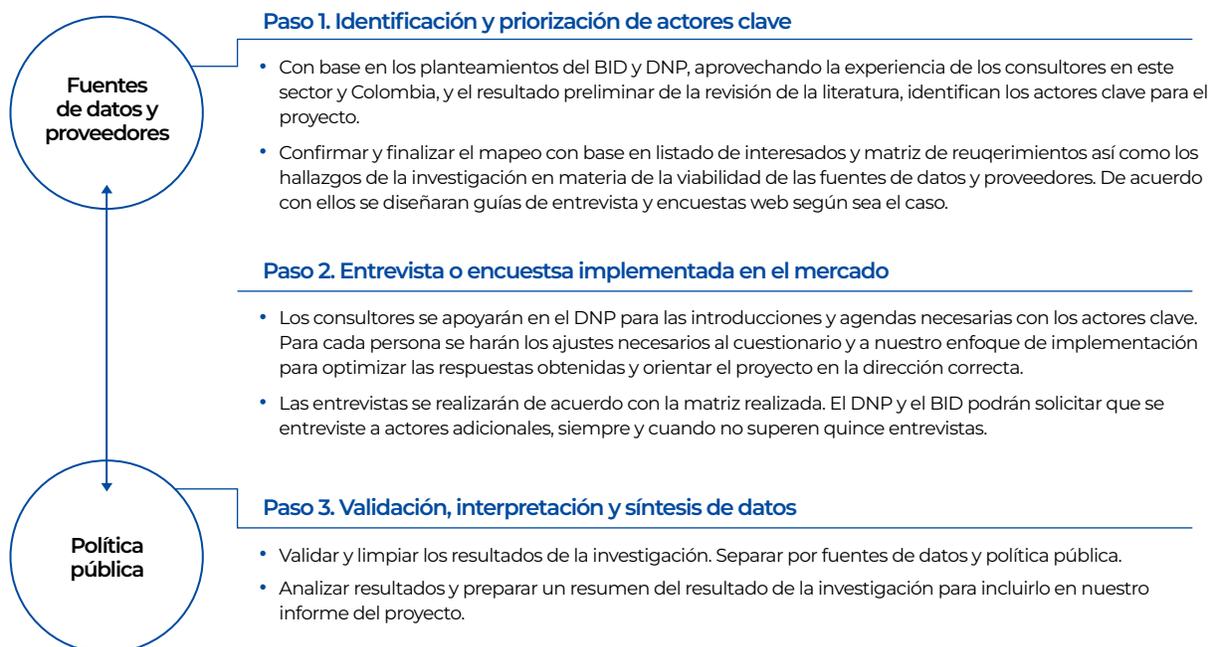
El diseño de la investigación con actores claves partió de las discusiones y reuniones con los patrocinadores del proyecto, en donde se discutió la matriz de interesados. También, se tomó en cuenta la experiencia actual de los consultores, así como los hallazgos preliminares de la revisión de literatura (figura 3-6).

Las guías de entrevista tuvieron como objetivos:

- Comprender los requerimientos clave de las instituciones financieras para la solución usando fuentes de datos alternativos y como la banca pública de desarrollo lo puede subsanar.
- Comprender cómo empata con las barreras en la implementación de soluciones de *scoring* Alternativa Digital, a partir de las necesidades de las entidades financieras para colocación de crédito.
- Identificar cuáles son los hábitos de préstamo de las mipymes desde la perspectiva de las entidades financieras.
- Comprender cómo se está implementando la política pública actualmente, qué barreras hay, y cómo se visualiza la atención de las necesidades futuras desde una perspectiva de política pública.

Figura 3-6. Metodología para la ejecución y análisis de las entrevistas con actores clave

Investigación primaria - Entrevistas “*On demand*”: Máximo 15 entrevistas. Priorizar de acuerdo a plan de trabajo, producto de consultoría previendo disponibilidades en diciembre y enero.



Fuente: elaboración propia.

3.6. Análisis

Una vez que se reunieron todos los hallazgos, se filtraron y agruparon para conformar la estructura de la presente investigación (figura 3-7).

Figura 3-7. Filtrado de hallazgos



Fuente: elaboración propia.

3.7. Síntesis

A partir del análisis de todos los hallazgos, se conforma esta investigación siguiendo la siguiente estructura: la primera sección, describe la situación general de inclusión financiera en Colombia, con un enfoque en mipymes y generalidades de las fuentes de datos (incluyendo definiciones que se usarán a lo largo del documento); la segunda sección, explica la metodología y las herramientas usadas; la tercera parte (secciones 4 a la 10), se refiere a cada fuente usando los criterios antes descritos; la cuarta sección, detalla el entorno normativo de las fuentes de datos seleccionadas; la quinta, hace recomendaciones en cuanto a política pública y para el lanzamiento de un piloto, así como aspectos a considerar; la última sección, resume las conclusiones de la investigación.

3.8. Atributos técnicos y funcionales de evaluación

Como se comentó en la sección 3.1.1, la revisión de literatura se divide en dos partes: la primera, consistió en recabar información teórica que fundamenta el uso de cada fuente de datos; y la segunda, en profundizar respecto a las fuentes seleccionadas.

Para seleccionar las fuentes de datos detalladas más adelante, se estableció una matriz de evaluación basada en los atributos técnicos y funcionales (criterios de calidad de datos definidos en la sección 2.2), además de agregar criterios de facilidad de implementación (derivados de experiencias internacionales y de otras entidades en Colombia) y de cumplimiento normativo (tabla 3-1).

Tabla 3-1. Criterios de evaluación de fuentes alternativas para su uso en riesgo de crédito

Criterios de selección	Telefonía móvil (Telco / CDR)	Datos de dispositivos <i>smartphone</i>	Datos de redes sociales	Datos de <i>e-mail</i>	Datos de psicometría	Datos de <i>open banking</i>	Transacciones públicas
Calidad de los datos - Disponibilidad	✓	!	!	!	✓	!	!
Calidad de los datos - Cobertura	!	✓	!	!	✓	×	!
Calidad de los datos - Especificidad	!	!	×	!	✓	✓	!
Calidad de los datos - Ortogonalidad	!	✓	?	?	✓	✓	✓
Facilidad de implementación	×	!	×	×	!	✓	✓
Cumplimiento normativo	!	×	!	!	✓	!	!

Fuente: elaboración propia.

Como se muestra en la tabla 3-1, se calificó cada fuente de datos contra los criterios establecidos. Cada fuente fue calificada usando las siguientes escalas:

- ✓ Amplia en Colombia (p. ej. Disponibilidad alta de datos de psicometría).
- ! Viable, pero con retos/limitaciones en implementación.
- × No existe/Barreras altas de implementación.
- ? No hay información suficiente para evaluar el criterio.

El resultado de la evaluación muestra que las fuentes de datos más relevantes para mipymes en Colombia son datos de psicometría, datos de *open banking* y datos de transacciones públicas.

Como se detallará en las secciones correspondientes, en la tabla 3-1 se muestra que la cobertura para *open banking* existe con barreras debido a los proveedores disponibles y madurez de sus soluciones. Así mismo, los datos de transacciones públicas son ortogonales y con una relativa facilidad de implementación, sin embargo, los otros criterios de calidad se cumplen de manera limitada.

A partir de la siguiente sección (3.6), la metodología abarcará únicamente las fuentes de datos seleccionadas (psicometría, *open banking* y transacciones públicas).

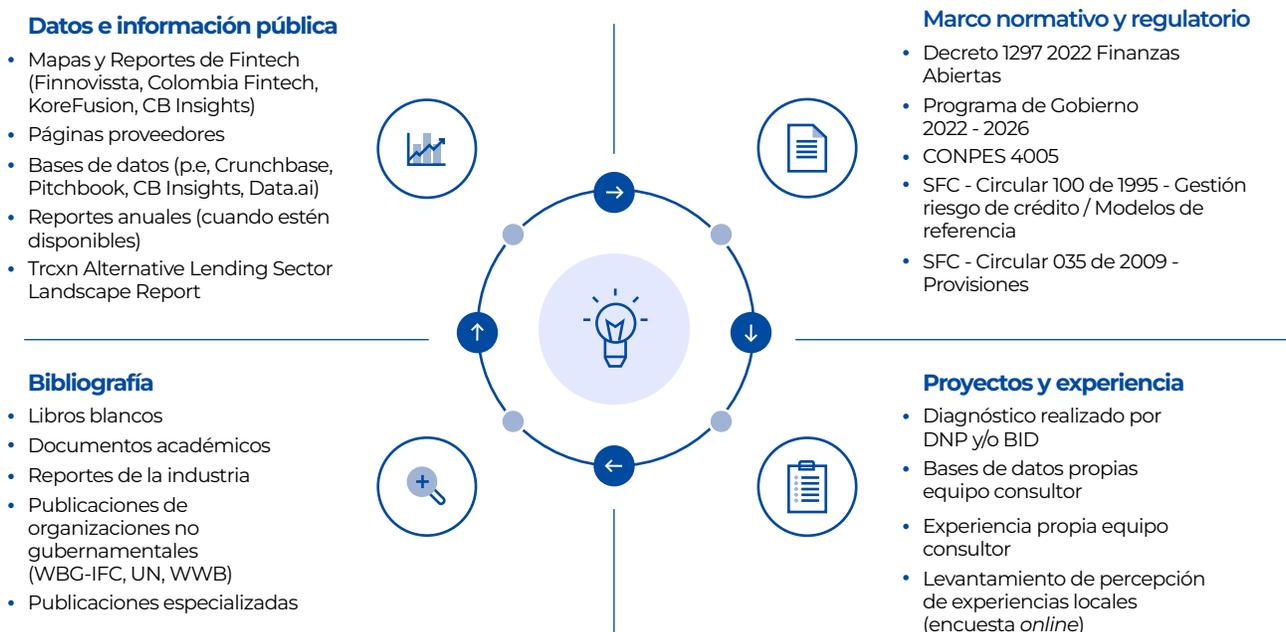
3.9. Herramientas y técnicas para levantamiento de información

Con el fin de llegar a conclusiones y recomendaciones adecuadas al alcance del proyecto, se determinó una serie de herramientas y técnicas para llevar a cabo una investigación rigurosa.

a. Fuentes de información

La revisión de literatura se realizó desde cuatro fuentes de información (figura 3-8).

Figura 3-8. Cuatro fuentes de información para revisión de literatura



Fuente: elaboración propia.

Esta estructura de la revisión de literatura recaba la información suficiente para evaluar cada fuente de datos con respecto a su calidad en el contexto de las mipymes en Colombia.

b. Gestión de interesados

Los interesados identificados se gestionaron mediante una matriz (esta herramienta permite tener una visibilidad concreta sobre el tipo de participante y la relación que puede tener con el proyecto).

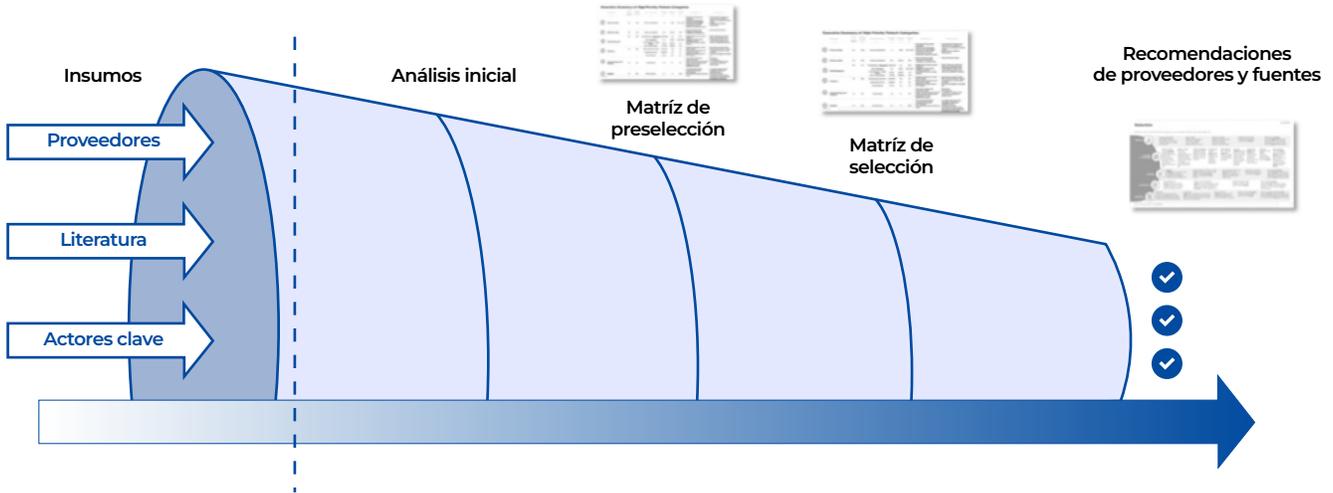
c. Preselección y selección de proveedores

Se diseñaron dos matrices para la conformación de recomendaciones de proveedores.

- Preselección
- Selección de proveedores

Estas matrices se basaron en los hallazgos preliminares y se usaron para evaluar a los proveedores identificados (figura 3-9).

Figura 3-9. Evaluación de proveedores

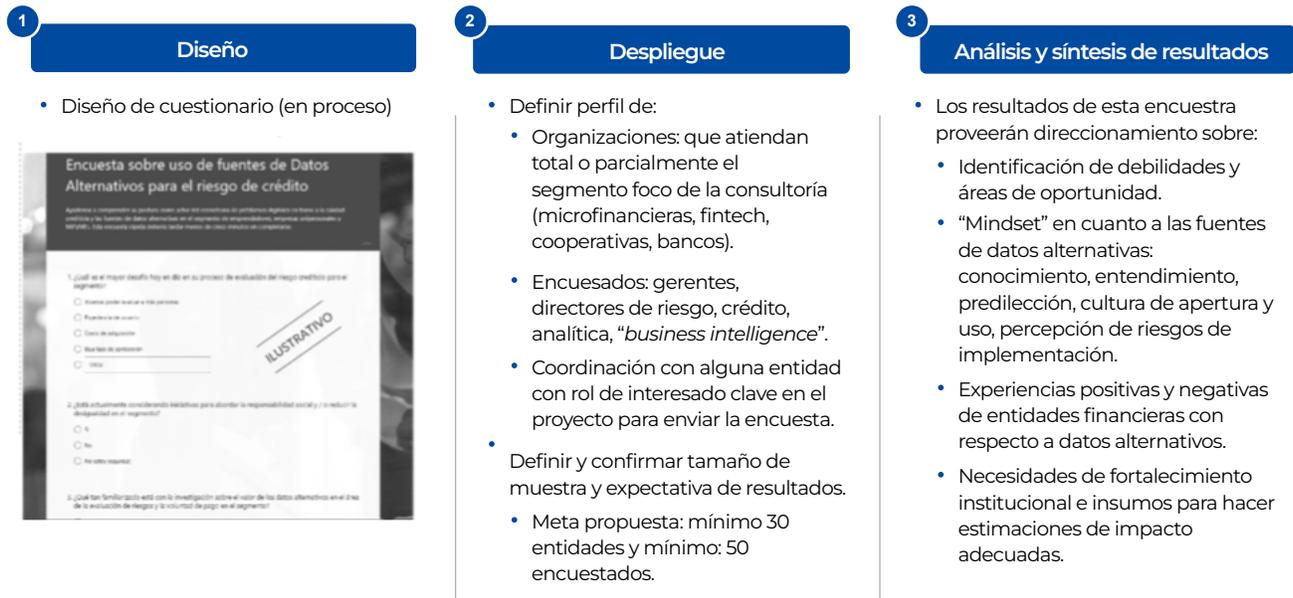


Fuente: elaboración propia.

d. Encuesta a instituciones financieras interesadas en el proyecto

Con el fin de entender las necesidades de las instituciones financieras objetivo de esta investigación, se realizó un cuestionario a manera de encuesta (figura 3-10). De esta forma se recabaron datos que sobre las necesidades y barreras de las entidades de cara a la implementación de soluciones de *scoring* alternativo. Los datos recabados fueron analizados y se conformaron las recomendaciones finales de la investigación.

Figura 3-10. Metodología para realizar encuestas a las instituciones financieras relevantes



Fuente: elaboración propia.

4. Datos de redes sociales

4.1. Generalidades en Colombia y en el mundo

4.1.1. Contexto y definición

Las redes sociales como fuente de datos alternativos parten del supuesto de que gran parte de la población no tiene un historial financiero establecido, pero tienen una presencia en línea a través de las redes sociales tanto en interacciones en sus teléfonos inteligentes como en computadores personales.

Las redes sociales son populares entre las pymes como herramientas para presentarse al público y adquirir nuevos clientes: plataformas como LinkedIn y Facebook recopilan perfiles de mipymes y reportan información sobre sus negocios, sus iniciativas más recientes, así como sus interacciones con sus clientes. Se podría argumentar que los prestadores de mipymes pueden acceder a esta fuente alternativa de datos utilizando las API proporcionadas por las plataformas, mediante agregadores y agencias/firmas de analítica de datos y utilizar esta información para crear un modelo de calificación crediticia más completo que sea transparente, rápido y barato y, por lo tanto, adecuado para las pymes.

4.1.2. Estudios y experiencias internacionales

Distintas instituciones y empresas han analizado el papel de las redes sociales para el acceso a crédito de las mipymes. Un estudio llevado a cabo en USA y Turquía encontró que el número de Me gusta y Seguidores, así como la riqueza del contenido, la interacción con los clientes y el uso del lenguaje en Facebook y Twitter, son indicadores que pueden ser usados para medir el desempeño de un negocio.

En otro caso, se estudió Vkontakte⁶ para la calificación crediticia de las microfinanzas. Se usaron los datos para discriminar entre prestatarios cumplidos e incumplidos en préstamos personales. La investigación mostró que los datos de las redes sociales pueden enriquecer efectivamente el score crediticio clásico. También, que la incorporación de métricas de redes sociales puede mejorar la predicción de repago en un 18 %.

En China se construyó un modelo de score crediticio para consumidores, fusionando información de redes sociales recopilada de la plataforma de préstamos (persona a persona) PPDa⁷ con la información que se usa en el modelo tradicional de calificación crediticia. El modelo demostró ser capaz de detectar un 86 % de clientes morosos con una precisión general del 85 %.

⁶ Vkontakte es una red social rusa con más de un millón de usuarios activos y más popular en Rusia.

⁷ PPDai (Shanghai Paipaidai Financial Information Service), se fundó en 2007 y tiene su sede en Shanghai. PPDai.com es la primera plataforma en línea en China que ofrece préstamos sin garantía P2P (*peer-to-peer*).

En Filipinas, México y Colombia, la Fintech Lenddo desarrolló un algoritmo que agrega datos de cuentas de redes sociales tales como amigos, frecuencia de interacción, intereses, mensajería e historial de navegación (Harvard Business School Alumni, 2016). Este algoritmo se mantuvo mediante el mapeo de relaciones entre más de 120 perfiles de redes sociales en Facebook, correo electrónico, LinkedIn, Twitter y Yahoo.

4.2. Proceso de recolección o producción

Los datos de las redes sociales brindan cantidades de información que pueden respaldar las decisiones sobre la calidad crediticia de un solicitante. Investigaciones anteriores, así como experiencia del equipo consultor muestran que datos de redes sociales en la calificación crediticia han sido utilizados en modelos de crédito en el pasado, sin embargo, estos modelos se limitaron a evaluar préstamos personales o de consumo. Para tener más contexto sobre las implicaciones del uso de datos de redes sociales referirse al punto 1.1: Entorno normativo.

Teniendo en cuenta un esquema de trabajo para análisis predictivo con datos alternativos, las variables (“variables estadísticas”) construidas a partir de datos de redes sociales se agrupan en tres categorías: (a) relacionadas con el volumen (p. ej. número de *tweets* o “me gusta”); (b) relacionadas con el sentimiento; y (c) características del perfil de los usuarios (p. ej. número de seguidores y la ubicación). La tercera etapa es la creación de modelos predictivos, que incluye la selección de un conjunto de métodos utilizados en el modelado, combinando los predictores con otros que no son de esa red social o de esa fuente de datos y aplicando técnicas estadísticas para construir un modelo de predicción. Finalmente, la evaluación del modelo incluye la elección de las métricas de evaluación y la definición de la especificación del modelo de referencia para comparar.

Algunas de las variables obtenidas a partir datos de redes sociales que se han propuesto en distintos estudios están compilados en la tabla 4-1.

Tabla 4-1. Compilación de distintos objetivos y variables usadas por red social para predecir comportamientos de mipymes

Objetivo	Plataforma	VARIABLES PRINCIPALES
Construir un modelo de calificación crediticia para préstamos personales	Vkontakte (Rusia)	<ul style="list-style-type: none"> • Afiliación política • Cantidad de fotos o videos publicados • Número de días desde la última visita • Número de días desde la última publicación
Construir un modelo de calificación crediticia para microcréditos	Facebook	<ul style="list-style-type: none"> • Información demográfica • Página seguida y preferencia • Un conjunto de métricas de red
Construir un modelo de calificación crediticia para préstamos personales	PPDAI (China)	<ul style="list-style-type: none"> • Cantidad de <i>followers</i> en la plataforma • Puntuación de contribución y reputación • Puntuación del grupo al que pertenecía
Desarrollar KPI y métricas para evaluar el uso de las redes sociales en la promoción	Facebook, Twitter	<ul style="list-style-type: none"> • Volumen y tipo de retroalimentación • Número de vistas, comentarios, compartidos • Número y tasa de crecimiento de fans • Calificaciones, número de me gusta o no me gusta • Número de <i>retweets</i>, menciones

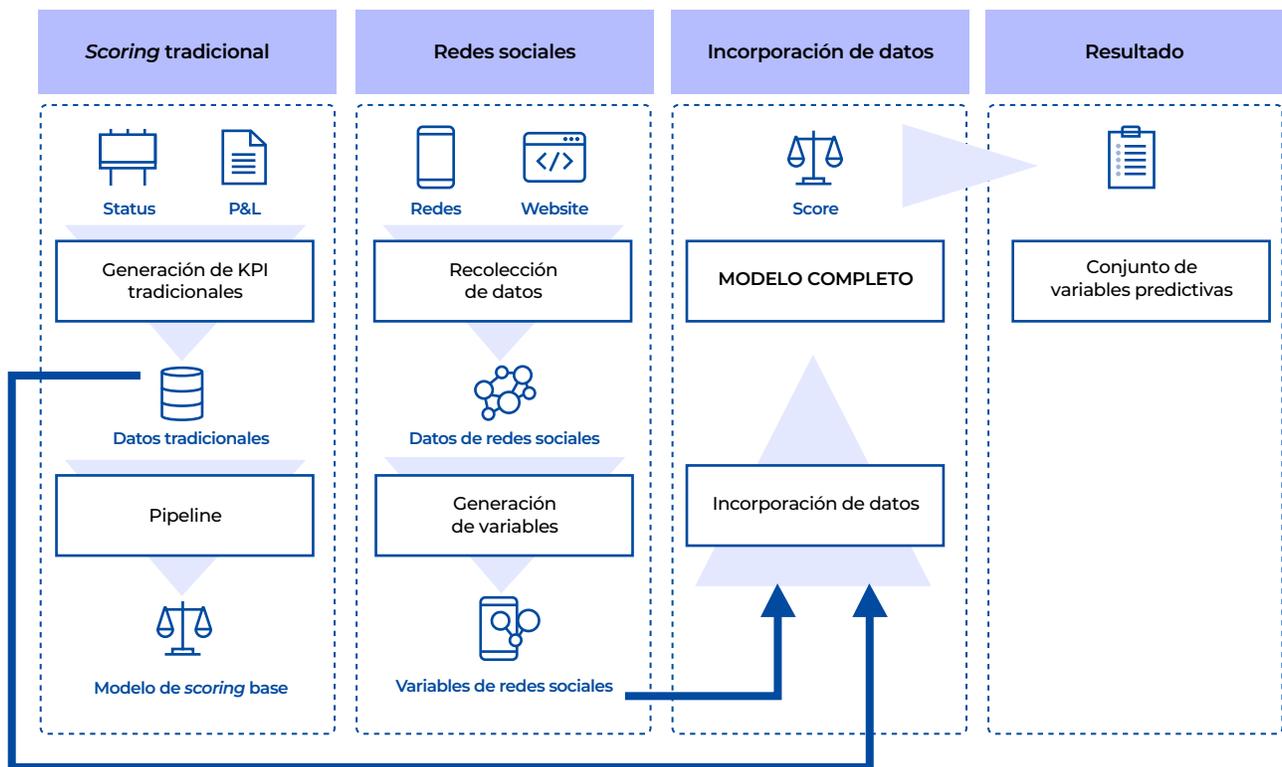
Tabla 4-1. Compilación de distintos objetivos y variables usadas por red social para predecir comportamientos de mipymes (continuación)

Objetivo	Plataforma	VARIABLES PRINCIPALES
Desarrollar métricas para evaluar el compromiso y el estado de ánimo de las partes interesadas de una página corporativa de Facebook	Facebook	<ul style="list-style-type: none"> • Porcentaje de publicaciones comentadas, compartidas y gustadas • Promedio de me gusta, compartidos, comentarios • Promedio de me gusta, compartidos, comentarios por relación de fans • Comentario positivo y negativo
Predecir el valor de la empresa	CNET, Lexis Nexis	<ul style="list-style-type: none"> • Volumen y sentimiento de las reseñas • Volumen y nivel de <i>rating</i>
Desarrollar métricas para evaluar el <i>marketing</i> de las PYMES.	Facebook, Twitter	<ul style="list-style-type: none"> • Número de fans o seguidores • Frecuencia de actualizaciones • Riqueza y relatividad del contenido • Interacciones

Fuente: elaboración propia con información de las bibliografías mencionadas.

De esta manera, con el esquema de trabajo antes mencionado y con las variables, el score tradicional se complementa con información no financiera como se muestra en la figura 4-1.

Figura 4-1. Incorporación de redes sociales en modelo de scoring orientado a mipymes



Fuente: elaboración propia con información de las bibliografías mencionadas.

4.3. Conversión, fricción y barreras de adopción en el proceso

Aunque los datos de las redes sociales se han utilizado ampliamente en la calificación crediticia para préstamos personales o de consumo, pocas veces se han empleado en el contexto de las pymes comercialmente y fuera de estudios académicos.

A pesar de su riqueza intrínseca, el uso de la información de las redes sociales para la calificación crediticia presenta desafíos (Bocconi, 2015):

- Los datos no son accesibles fácilmente, a menos que los proporcione directamente la empresa que se está evaluando, lo que no siempre es el caso; por lo tanto, se deben diseñar técnicas inteligentes de recopilación de datos para su adquisición.
- Utilizar los datos de las redes sociales para la calificación crediticia de mipymes es una práctica naciente. Varias empresas ofrecen servicios comerciales de evaluación crediticia (parcialmente) basados en las redes sociales existentes (por ejemplo, Lenddo o Kabbage), pero la naturaleza y el rendimiento de sus métodos se ocultan por razones comerciales. Actualmente, las mejores prácticas sobre cómo usar los datos de las redes sociales en la calificación crediticia, hasta la fecha, aún no se han establecido.

Los datos de redes sociales, como Facebook, por ejemplo, pueden resultar relevantes para identificar voluntad de pago, sin embargo, no son necesariamente precisos para medir la solvencia o capacidad de pago, ya que solo pueden hacer una “inferencia” acerca del comportamiento financiero, los ingresos y hábitos de consumo, pero poca información relevante sobre las deudas o el historial de pagos de los usuarios.

Una consideración importante para usar datos de redes sociales radica en que usuarios pueden crear perfiles falsos, publicar información engañosa o manipular su reputación en línea para elaborar una “huella digital artificial”, es decir, una huella digital que no es resultado natural del uso habitual de la misma red social y esto puede influir en los resultados que un modelo de score de crédito arroje, a esta práctica se le puede llamar “*informed gaming*” o “hacer trampa de manera informada” en redes sociales.

Adicional a lo anterior, acontecimientos como los ocurridos en el caso Cambridge Analytica⁸; donde como consecuencia de dicho escándalo se produjeron multas y cambios normativos en el uso de datos de redes sociales, que condujeron a la implementación de políticas de privacidad de las distintas redes sociales, lo que por un lado protege de mejor manera a los usuarios, pero también restringió el uso de subconjuntos de datos a los proveedores que usan las API (glosario de términos) de dichas redes sociales, disminuyendo significativamente funcionalidad, poder predictivo y en generar el valor percibido de dichas fuentes de datos alternativos en los procesos de análisis de crédito.

⁸ Cambridge Analytica fue una firma de consultoría política que utilizó datos de Facebook de manera ilegal para influir en las elecciones presidenciales de EE. UU., en 2016.

5. Datos de psicometría

5.1. Generalidades en Colombia y en el mundo

La evaluación psicométrica es un enfoque científico dentro del campo de la Psicología, que se utiliza para medir el carácter o las capacidades de un individuo. Las evaluaciones psicométricas se han utilizado en varios dominios durante muchos años y en el contexto de la evaluación crediticia, las pruebas o evaluaciones psicométricas se pueden utilizar para comprender la voluntad y la habilidad de pago de un individuo. El enfoque en la voluntad de pago determina si el solicitante tiene la propensión a honrar sus deudas, es decir, si en general “quiere pagar” y explica la “irracionalidad” de las personas en cuanto a su decisión de pago.

Un reto que existe al hacer análisis de factibilidad para determinar si se usará la psicometría en el proceso de crédito es la malinterpretación de esta metodología. La psicometría no sustituye el análisis de capacidad de pago, es decir una evaluación psicométrica no determina si un individuo posee capacidad de pago; sino que ofrece un conjunto de datos adicionales para identificar la ‘voluntad’; independiente a contar con los recursos para pagar en el futuro (capacidad de pago) o si lo ha hecho en el pasado (historial de crédito).

Con base en lo anterior, mientras que la evaluación de crédito tradicional considera la capacidad de pago de un solicitante al observar la información histórica de pago, una puntuación de crédito alternativa que incluya psicometría, indicará la voluntad de pago de un solicitante al considerar *insights* de su carácter, personalidad, comportamiento y capacidades.

5.2. Fundamento teórico/académico

La psicometría se fundamenta en dos grandes corrientes teóricas y metodológicas: las ciencias conductuales y el análisis de la personalidad.

5.2.1. Ciencias conductuales

El término *behavior credit*, (calificación crediticia) tradicionalmente se ha referido a las metodologías sobre cómo calificar a clientes que ya están en nuestro portafolio con base en su comportamiento de pago. Evidentemente este representa retos distintos a los de calificar a un prospecto nuevo como se hace con los *scores* de adquisición. No obstante, dos grandes situaciones están ampliando el alcance de esta temática.

- a. La crisis derivada de la emergencia sanitaria que obligó a replantear los modelos tanto de adquisición como de mantenimiento ante la falla sistémica de mercado al tener una súbita alza de cartera morosa.
- b. La creciente tendencia del uso de *behavioral economics* o economía conductual en todos los ámbitos.

Esto ha abierto la conversación a identificar nuevas tendencias, metodologías y tecnologías para evaluar y entender de una manera más profunda a prospectos y clientes actuales y ampliar el alcance de esa palabra “*behavior*”, no solo en cuanto a comportamiento de un cliente actual en el portafolio sino a evaluar el “comportamiento conductual” de ese solicitante como persona o como empresario.

Gran parte del estudio de las ciencias conductuales en el campo de acceso y uso de productos financieros se ha fundamentado en la teoría de “economía conductual” o *Economía Conductual*, la cual es una antítesis a la teoría económica neoclásica. Por un lado, la teoría económica neoclásica tiene como principales principios que las decisiones económicas son: racionales, buscan la maximización de la utilidad, asume que los individuos tienen preferencias consistentes, y considera obtener el máximo beneficio de satisfacción de estas preferencias; poner recursos para lograr la máxima eficiencia es el eje de este pensamiento donde para este consumidor individual el término “utilidad”, se define como la máxima satisfacción, felicidad o bienestar que un bien o servicio le puede ofrecer (Corr & Plagnol, 2018).

Sin embargo, en la práctica, el humano no es racional dentro de sus decisiones en el día tras día, por ejemplo, la racionalidad de la teoría económica neoclásica dice que los costos hundidos nunca se consideran en la toma de decisión, puesto que como su nombre lo dice ya se ejecutaron, sin embargo, la psicología de las personas siempre lo tienen en cuenta (Corr y Plagnol, 2018).

La economía conductual utiliza conceptos como *delayed gratification* o “gratificación aplazada”; el ser humano no solo es irracional, sino que además la gran mayoría no tiene conocimientos estadísticos ni financieros básicos; sin embargo, la teoría económica neoclásica considera que el grueso de la población sí tiene estos conocimientos. Al realizar un ejercicio de gratificación aplazada a través de una prueba psicométrica, se pueden identificar diferencias en cómo funciona el cerebro dando lugar a diferencias en personalidad, inteligencia, así como gustos y preferencias, y determinan como se responde a eventos externos como, por ejemplo, reacción ante incentivos económicos, impulsividad al hacer una compra o decisión de honrar el pago de una deuda (Corr y Plagnol, 2018).

El uso de las teorías de economía conductual en la industria financiera ha sido a través de técnicas como el *nudging*, que consiste en entender que el ser humano tiene un comportamiento extraordinariamente complejo, tanto física como mentalmente, pero es muy beneficioso reconocer que el cerebro humano y el comportamiento pueden fallar sistemáticamente en tareas específicas. Los trucos visuales, como adivinar el tamaño de dos tablas dibujadas en diferentes orientaciones en el espacio que parecen distintas pero que de hecho son iguales, posibilita la comprensión de que el error humano en elección y determinación es algo muy real y fascinante a considerar en especial en decisiones de crédito.

Ahí es donde radica la importancia de la psicología al comprender cómo los errores humanos afectan la visión tradicional de la economía, creando nuevos paradigmas para evaluar cómo las personas toman decisiones. Los humanos cometen errores con regularidad, basados en falacias y otros factores involucrados en la toma de decisiones; los autores proponen que un *nudge* o (empujón) puede influenciar a las personas a tomar decisiones y que haciéndolo responsablemente se considerarían mejores decisiones.

Un ejemplo de esto se ve en productos financieros de ahorro bloqueado o ahorro de única destinación. Son productos que generalmente no ofrecen liquidez, representan un alto costo transaccional y no pagan interés. La principal característica se refiere a que el dinero no pueda ser retirado para propiciar una disciplina de ahorro “forzado”. Un producto de esta naturaleza desde el punto de vista de la teoría económica neoclásica se vería que es una desventaja y no una ventaja, sin embargo, el ser humano por su irracionalidad en ocasiones necesita barreras para resistir esa “tentación”. La creación e implementación de productos como este, usó los principios del *Mental Accounting* o contabilidad mental.

Contabilidad mental es un sistema de control natural del cerebro para poner barreras a la tentación ya sea evitando o reduciendo estas (Thaler y Sunstein, 2021); para este caso, incentivar el ahorro se puede hacer sobre objetivos específicos. Anteriormente se utilizaban métodos más tradicionales como los “jarrones”, cuyo principio se ha tecnificado por ejemplo con billeteras móviles. La cuenta en línea Wise⁹ como solución de ahorro tiene una funcionalidad de “jarrones” para separar el dinero en destinaciones específicas.

Otro caso es Nequi¹⁰, que entre otras funcionalidades utiliza el concepto de “colchones” para ahorros grandes y opciones explícitas para evitar el retiro de esos recursos, por ejemplo tiempo (configurar una duración específica para cuánto tiempo deberá estar ese dinero guardado), cuidado de un amigo (configurar a un amigo usuario también de la plataforma, el cual se pone al cuidado de ese ahorro a manera de “guardián”, que deberá autorizar cuando el dinero sea retirado) y trivia (preguntas de seguridad que dificultan el retiro del dinero y obligan al usuario a “reconsiderar” su decisión de dejar de ahorrar). Al aplicar este principio a productos de crédito, se ha demostrado que individuos que llevan esta contabilidad no muestran aversión a la deuda y, por lo tanto, se vuelven susceptibles a los riesgos asociados con este estilo de gestión financiera, que en casos extremos puede conducir al sobreendeudamiento (Swacha-Lech y Solarz, 2019).

5.2.2. Análisis de la personalidad

Las personas tienen una disposición de personalidad perdurable que predice en parte lo que harán y hay una ciencia detrás de la personalidad en cuanto a cómo se mide la personalidad y qué puede predecir (Nettle, 2020).

⁹ <https://wise.com/> cuenta electrónica que permite enviar dinero, recibir pagos y gastar dinero de forma internacional.

¹⁰ <https://www.nequi.com.co/> en Colombia y Panamá opera como plataforma financiera en un cien por ciento digital.

El análisis de la personalidad se refiere a un conjunto de conceptos, cuyo uso es una evidencia sólida en la que los psicólogos están de acuerdo. Este conjunto de conceptos se denomina modelo de cinco factores. Este modelo ha emergido de una extensa investigación siendo hoy en día el marco más completo, fiable y útil para discutir la personalidad humana que jamás haya tenido un psicólogo. Hay cinco dimensiones principales a lo largo de las cuales varían todos los caracteres humanos y dicen mucho sobre las formas en que pueden comportarse en sus vidas.

La personalidad afecta las muy particulares maneras de pensar, sentir y comportarse en las diferentes situaciones y en el transcurso del tiempo. Hay una interrelación entre una situación en particular y los factores de personalidad, ya que el cómo se reacciona depende en gran medida de la propia personalidad. Es decir, diferentes comportamientos ante situaciones similares son explicado por las diferencias de personalidad (Corr y Plagnol, 2018); es así como la razón principal de estos comportamientos es por genética y fisiología como resultado de cómo funciona la mente.

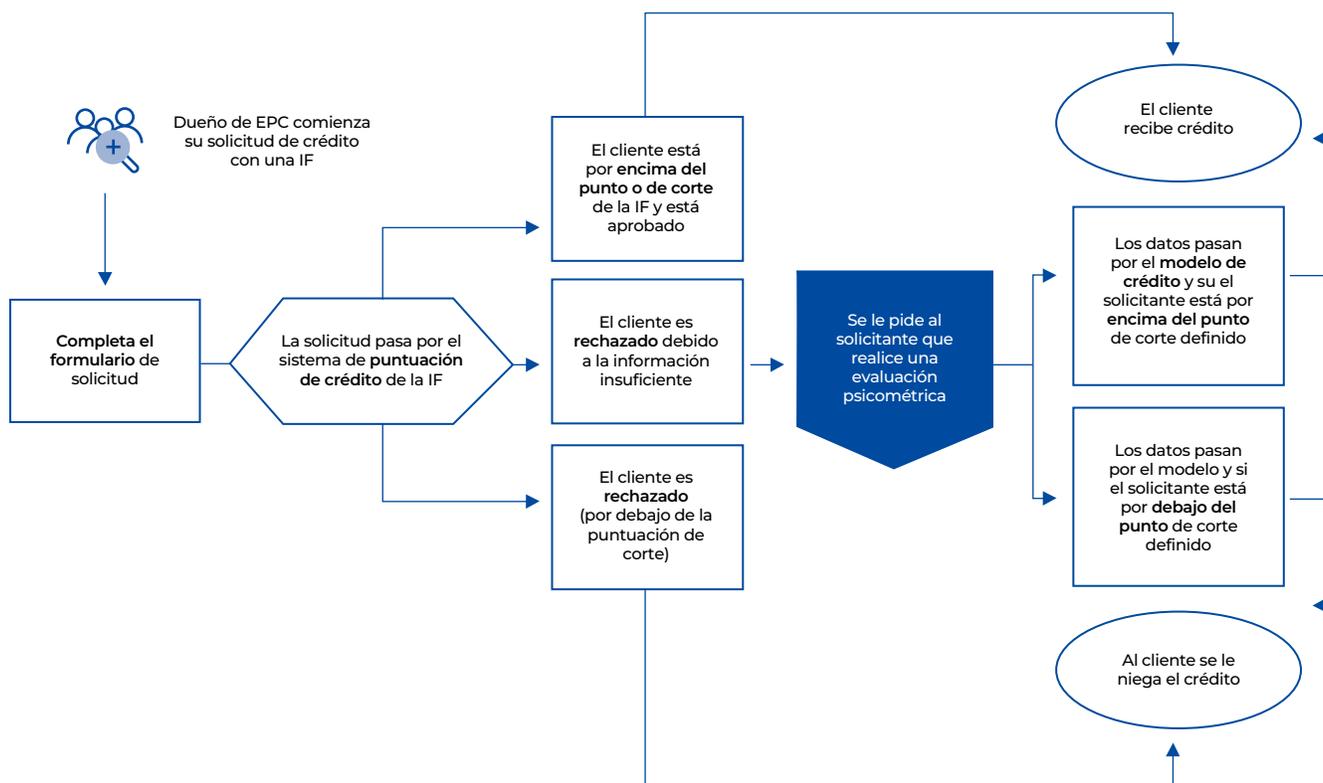
Se debe asumir que cualquier influencia social debe pasar por el cerebro para que ejerza cualquier tipo de influencia. Esas diferencias en cómo funciona el cerebro dan lugar a cómo lo hace la personalidad, inteligencia, así como gustos y preferencias, que determinan la forma cómo se responde a eventos externos (Corr y Plagnol, 2018). Lo crucial es que estas diferencias entre los distintos factores o dimensiones de la personalidad afectan la percepción del entorno que se estructura en el cerebro o mente y esa percepción también puede afectar el resultado traduciéndose en una probabilidad de incumplimiento o de impago.

Los modelos de score crediticio entonces buscan es llevar la teoría psicológica de la personalidad y esa teoría económica es comportamental al ámbito de riesgo crediticio a través de un modelo estadísticamente predictivo sobre la probabilidad de impago de una mipyme.

5.3. Proceso de recolección o producción de datos psicométricos

Las IF están renunciando a una gran oportunidad al no prestar a los clientes rechazados por causales blandas, principalmente debido a la información insuficiente para evaluar su calidad crediticia. Ante este caso de uso, la solución sería “reevaluar” o “rescatar” esa solicitud de crédito. Generalmente se hace en un proceso similar al que se presenta en la figura 5-1.

Figura 5-1. Flujo de proceso de recolección de datos psicométricos en un score de rescate



Fuente: tomado y adaptado de <https://lenddoefl.com/>

5.4. Disponibilidad, cobertura y veracidad de datos

Como se comentó anteriormente, uno de los mayores beneficios de la psicometría es su cobertura. Algunos proveedores inclusive llegan a afirmar que la evaluación psicométrica es el único score de crédito universal (Begini, 2022). Esto opera bajo el fundamento de que todos los individuos tienen una personalidad, lo que significa que con las herramientas adecuadas se evalúa a cualquier persona.

Una ventaja de cobertura y disponibilidad se ha identificado en las fuentes de datos psicométricas al utilizar un cuerpo de conocimiento referente al amplio estudio de la personalidad y comportamiento, de parte de la Psicología han sido en su mayoría construidas con fundamentos de aplicación global. Esto permite que, por ejemplo, un modelo utilizado en un segmento y producto financiero similares en África¹¹ pueda aplicarse por primera vez como un modelo genérico en Perú.

¹¹ Ejemplo de piloto en Perú. Fue la primera implementación de una solución psicométrica con un score de crédito de África basado en información de 920 observaciones con datos de reembolso de préstamos (Arráiz, Bruhn, Ruiz, y Stucchi, 2018).

5.5. Requerimientos para la implementación¹²

La Implementación de las soluciones de psicometría requiere del desarrollo de un proyecto en distintas etapas, a continuación, se describen algunos de los requerimientos de las principales:

5.5.1. Creación o configuración de la evaluación psicométrica para la IF

Consiste en el desarrollo o configuración de una evaluación psicométrica que por lo general opera en una web app. Dependiendo del proveedor esta evaluación deberá ser configurada en el idioma, con los colores y logo institucionales de la IF. Según del producto, segmento y geografía esta evaluación tendrá algunas redacciones específicas o ejercicios específicos.

Por lo general, esta web app se aloja como una URL independiente y la IF puede invitar a los solicitantes de crédito a realizar y completar la evaluación.

5.5.2. Integración técnica

En esta etapa un especialista en integración del proveedor da soporte al equipo técnico del cliente sobre la creación de evaluaciones y alojamiento de la URL, independiente sobre la cual se realizará la evaluación psicométrica y la integración de API o Webhook para que el cliente pueda consumir las puntuaciones, *insight* (perspectivas), variables (también llamada “variable estadística”) y demás datos de resultado que se hayan definido en el alcance del proyecto.

5.5.3. Recolección de datos

Posterior a la integración, el cliente está en capacidad de lanzar la evaluación psicométrica y de esta manera recopilar datos psicométricos prospectos reales. De la misma manera, resulta de gran valor obtener datos de repago cómo estos prospectos, en el evento en que se conviertan en clientes al haberse abierto un crédito, van pagando su crédito con la IF originadora.

La etapa de recolección de datos depende de los volúmenes de solicitud y desembolsos que propicie la IF, así como los tiempos de vencimiento del crédito. Es importante denotar que cuantos más datos representativos se puedan recopilar sobre los pagadores “buenos” y “malos”, mejor será el modelo que los equipos de ciencia de datos del proveedor o de la IF podrá construir.

5.5.4. Construcción o recalibración del modelo de score psicométrico

En esta etapa se realizan procesos iterativos de extracción de datos, limpieza de datos, procesamiento de los datos. Esto aplica tanto para los datos psicométricos como los de desempeño de repago de los créditos.

Los principales requerimientos asociados a esta etapa son:

¹² Compilación propia resultado de la revisión de los sitios web de documentación de negocio, técnica y analítica de los principales proveedores analizados descritos en detalle en el apéndice.

- a. Definición de altura de mora máxima a partir de la cual una operación de crédito puede ser considerada como “mala”.
- b. Selección de variables (también llamada “variable estadística”): se seleccionan las variables que serán usadas en el modelo para el ejercicio de entrenamiento de este. La potencia predictiva, la distribución y la estabilidad a lo largo del tiempo se verifican para cada variable. También se revisa la correlación entre ellos y con las definiciones predeterminadas. Como resultado, se clasifican y se crean diferentes grupos de variables.
- c. Selección de modelos: una vez creados los modelos preliminares, estos se aplican a las muestras de entrenamiento para seleccionar a los candidatos finales, que luego serán evaluados en las muestras de prueba y validación. En este paso se elaboran y evalúan diferentes indicadores de desempeño de modelos, como por ejemplo AUC, KS (indicador de Kolmogorov-Smirnoff) y Gini. También se consideran las características de entrada de los modelos, como los pesos de las variables y el PSI.

5.5.5. Implementación, despliegue y activación del modelo

Una vez entendiendo el desempeño esperado, puntos de corte y definiciones en la política de crédito este se implementa en la plataforma del proveedor de psicometría. Esto significa que una vez que los clientes solicitan un préstamo son conducidos a que realicen la evaluación psicométrica del proveedor, este recolecta los datos, los procesa y la IF recibe los puntajes dentro del ANS establecido para ser utilizados con fines de toma de decisiones.

Se deben realizar actividades de desarrollo y pruebas, tanto unitarias como integrales. Algunas de estas son incrustación de las variables en el modelo, pruebas de calificación del modelo, pruebas *Black box Testing-Dev/QA*, prueba de lanzamiento del modelo en producción, configuración de las URL (Uniform Resource Locators, por sus siglas en inglés de localizador uniforme de recursos) y credenciales del proveedor dentro del ambiente digital de la IF y pruebas de aceptación del usuario.

Se considera finalizada esta etapa, cuando se obtienen resultados de usuarios reales en un ambiente de producción, estos resultados pueden ser puntajes, *insights* o variables de acuerdo con el alcance acordado con el proveedor.

5.5.6. Monitoreo

Una vez que se implementa y activa un modelo, comienza un proceso periódico de su monitoreo. Según el proveedor se entre tres a seis veces al año, se recopilan y actualizan los datos, tanto con los que se construyeron las variables como los de desempeño de repago, para evaluar el desempeño del modelo y la estabilidad de la población. Estos resultados se inspeccionan y estudian con el doble propósito de garantizar que el modelo funcione como se espera y recopilar la información necesaria para la toma de decisiones basadas en las evidencias.

5.6. Resumen de beneficios potenciales

Uno de los principales beneficios de los datos psicométricos proviene del fundamento de que se obtienen a partir de la personalidad y todos los individuos tienen una, entonces tienen una disponibilidad y cobertura muy cercana a cien por ciento.

Existen resultados concluyentes sobre un piloto realizado a propietarios de pymes con el acompañamiento de IDB y uno de los cinco bancos más grandes de Perú (Arráiz, Bruhn, Ruiz, y Stucchi, 2018) donde se pueden resaltar los siguientes aspectos:

- Los resultados sugieren que la calificación crediticia psicométrica es un método de selección viable para los solicitantes de préstamos que no tienen historial crediticio.
- La herramienta psicométrica se aplicó a todos los clientes, independientemente de su historial crediticio, también destaca el poder de la información densa de los burós de crédito. Es decir, cuando existe mala información crediticia, parece más beneficioso para las instituciones financieras basarse únicamente en esta información en sus decisiones crediticias.
- Se concluye que la herramienta psicométrica permite que las IF ofrezcan crédito a las pymes que no tienen un historial crediticio y que fueron rechazadas con base en sus puntajes crediticios tradicionales, sin producir más incumplimiento.
- El aumento en el uso de préstamos para pymes resultó particularmente grande para los solicitantes de archivo delgado (definición de “archivo delgado”, glosario de términos).
- Para los solicitantes con archivos gruesos con información de buró, la evaluación psicométrica no condujo a un mayor acceso que no fuera con el banco asociado y sí a una reducción estadísticamente significativa en el puntaje crediticio de Equifax¹³, lo que puede deberse en parte a un peor comportamiento de pago.
- Se concluyó que la herramienta psicométrica puede aumentar el uso de préstamos para pyme en el corto plazo.

De acuerdo con el análisis realizado a proveedores con respecto a tiempos de respuesta para obtener un resultado, ya sean estos datos crudos, *insights* o *scores*, se identificó que dos de los principales proveedores aseguran que estos se obtienen inmediatamente; otros dos no indican tiempo ni promesa de valor, sin embargo, establecen que no hay demoras una vez el solicitante finalizó la evaluación psicométrica y el proveedor recolectó los datos satisfactoriamente¹⁴.

¹³ Equifax es una central de información con activos de datos, tecnología y capacidades analíticas que presta servicios de buró de crédito, defensor del consumidor y promoción de la educación financiera.

¹⁴ LenddoEFL (<https://lenddoefl.com/scoring>). Begin (<https://www.begin.co/psychometrics/>) Coremetrix: <https://demo.coremetrix.com/dynamic-feed-back>) al intentar realizar en la práctica el “cuestionario Coremetrix” Quash (<https://quash.ai/>)

5.7. Resumen de retos potenciales

Las evaluaciones psicométricas tienen un reto en cuanto a calidad de datos y es susceptibilidad de que la prueba sea falseada o que los individuos “hagan trampa” al contestar. A esto se le llama, *informed gaming* o “trampa informada”, que son básicamente estrategias para manipular los resultados de las pruebas psicométricas o de personalidad; las estrategias de juego informadas implican investigar el formato, el contenido y el sistema de calificación de la prueba o evaluación, y luego responder las preguntas de una manera que presente la imagen o el perfil deseado. Por ejemplo, alguien que quiere parecer más extrovertido puede elegir respuestas que indiquen alta sociabilidad, asertividad y entusiasmo, lo que puede conducir a resultados inexactos o engañosos.

La solución o herramientas de recolección de estos datos debe contener elementos que detecten o que prevengan que las respuestas sean falseadas recordando que no solo se tiene en cuentas las respuestas finales a la evaluación sino además el proceso y la forma de como el individuo responde o interactúa con la evaluación (LenddoEFL, 2018). Algunas consideraciones son:

- Tiempo y completitud: cuánto tarda el solicitante en realizar las preguntas, ejercicios, secciones o módulos.
- Patrones de respuestas: qué tan consistentes son las respuestas o resultados a lo largo de la evaluación.
- Intervención adecuada: el solicitante pone el suficiente grado de atención y tiempo a la realización de la evaluación.

En ocasiones los proveedores no pueden elaborar una puntuación, esto se debe principalmente a inconsistencias en los tres aspectos mencionados anteriormente, como, por ejemplo: respuestas inconsistentes o responder las preguntas de la prueba demasiado rápido. Ante lo cual el solicitante de crédito debería realizar nuevamente la evaluación o ser marcado como no calificable; ambos resultados muy seguramente derivarán en una mala experiencia de cara al usuario solicitante¹⁵.

Las evaluaciones psicométricas requieren que la información sea recolectada con una evaluación psicométrica al momento de la solicitud del crédito, es decir, en general con psicometría no se puede “adquirir” un conjunto de datos que contengan información psicométrica de solicitantes. Esto es una diferencia significativa con respecto a las otras fuentes de datos que puede representar un reto al momento de construir modelos de riesgo de crédito ya que no sería posible realizar *backtesting* (o pruebas retrospectivas, glosario de términos). El hecho de no poder realizar este tipo de pruebas retrospectivas si se desea utilizar un modelo genérico pueden aumentar la incertidumbre y las implicaciones en el VaR (*Value at Risk*, glosario de términos) para probar y entrenar un modelo.

¹⁵ Resultados obtenidos de <https://demo.coremetrix.com/dynamic-feedback> al intentar realizar en la práctica el “cuestionario Coremetrix”.
<https://coremetrix.com/es/nuestro-producto/>

Adicionalmente, el hecho de que los datos psicométricos sean recolectados al momento, dentro del flujo del proceso puede ser un reto en la coordinación para ciertos casos de uso y que además deriva en otro reto que es el tiempo de recolección de datos, las evaluaciones psicométricas pueden durar entre 10 y 45 minutos. En el caso del piloto de Perú, los solicitantes demoraron en promedio 45 minutos en completar la prueba, versiones más recientes de la herramienta tardan 25 minutos de acuerdo con el estudio (Arráiz, Bruhn, Ruiz, y Stucchi, 2018).

Un estudio interno de Te Creemos reveló las principales razones por las que las evaluaciones psicométricas no fueron aplicadas satisfactoriamente en sus sucursales: problemas con internet, alta rotación del personal, duración de la evaluación, falta de comprensión del solicitante entre otras referentes a entrenamiento del personal de la IF con respecto a los procedimientos internos (Rojas, 2014).

5.8. Principales proveedores

Se realizó un análisis de proveedores de datos de psicometría y de la conducta. Al ser esta fuente de datos un candidato a ser implementado, entender cuál es la oferta y su disponibilidad en el país será relevante para emitir las recomendaciones pertinentes de implementación. Los proveedores identificados se encuentran resumidos en la tabla 5-1, y las fichas que se desarrollaron se entregan mediante el Apéndice A.

Tabla 5-1. Preidentificación de proveedores psicometría

Compañía	HQ	Categoría	Sitio web	Año de fundación	Experiencia Colombia	Comentarios
Coremetrix ★	Reino Unido	Psicometría	coremetrix.com	1999	No	Empresa impulsada por Creditinfo, es el puntaje psicométrico para la evaluación de riesgos. Coremetrix es el nuevo nombre comercial de Visual DNA, luego de la adquisición en agosto de 2016.
LenddoEFL	Singapur	Psicometría y datos de dispositivos	lenddoefl.com	2010	Si	Ofrecen productos alternativos como <i>score</i> de crédito, verificación de identidad e <i>insights</i> para instituciones financieras.
Begini ★	Reino Unido	Psicometría y datos de dispositivos	begini.co	2021	Si	Análisis de datos de comportamiento para conocer y tomar decisiones sobre el riesgo crediticio, utilizando cero datos personales.
Quash	EUA	Psicometría y <i>open banking</i>	quash.ai	2018	No	Plataforma de motor predictivo que analiza data de préstamo y combina con datos demográficos y de comportamiento.
Harvesting	India	Psicometría y espacial	harvesting.co	2016	No	Plataforma de motor predictivo que analiza data de préstamo y combina con datos demográficos y de comportamiento.
CRIF India	India	Psicometría y metadatos de páginas web	crif.in	2005	No	La herramienta COSMO es un modelo de evaluación de riesgo crediticio, que estima rasgos de personalidad. También utiliza metadatos y datos de la solicitud de crédito.
Bemobi	Brasil	Conducta digital, dispositivos móviles y Telco	bemobi.com.br	2009	No	Gestión de riesgos, segmentación, seguridad e identidad, basados en datos alternativos móviles para mejorar los procesos comerciales en la industria financiera.

★ Proveedores preseleccionados.

Fuente: elaboración propia, basada en investigación previa, bases de datos y asociaciones Fintech

5.9. Análisis de propuestas comerciales

Inversión estimada

La inversión en soluciones de psicometría está asociada a las etapas del proyecto de implementación y requerimientos mencionados anteriormente. En la presente investigación se identificó que existen tres grandes rubros: instalación inicial, costo recurrente y servicios adicionales. A continuación, se describen cada uno de ellos con unos rangos de costo.

a. Instalación inicial

Por lo general es un costo único al principio de la relación con el proveedor, puede ser equiparable a Capex dentro de un proyecto y su costo puede variar principalmente debido al grado de complejidad del despliegue (por ejemplo, distintos productos, distintas geografías, distintos segmentos), fuentes de datos contratadas, duración del contrato y los volúmenes de transacciones esperados. Algunos proveedores ofrecen distintos escenarios de costos dependiendo de si se trata de una PoC o trial (glosario de términos), piloto o directamente un escenario de producción (tabla 5-2).

Tabla 5-2. Escenarios de costos de instalación inicial psicometría

	Trial o prueba de concepto	Piloto	Producción
Duración	Entre 1 mes y 3 meses	Entre 4 meses y 18 meses	+ 12 meses
Costo	Entre 1.000 y 5.000 USD	Entre 3.000 a 10.000 USD	Entre 5.000 a 40.000 USD

Fuente: elaboración propia, basada en investigación previa, bases de datos y estudios de cliente oculto.

Los servicios que generalmente están contemplados en esta instalación inicial incluyen el diseño de la evaluación, soporte a la integración y apoyo de un gerente de cuenta o un gerente de éxito de clientes para asesorar en el mejor uso de la solución.

b. Costo recurrente

Por lo general es un costo recurrente posterior a la implementación y lanzamiento de la solución, generalmente es mensual y está asociado a costos de licenciamiento de SaaS o costos transaccionales por consulta. Puede ser equiparable a Opex dentro de un proyecto y el costo puede variar principalmente debido al tipo de resultados contratados, fuentes de datos contratadas, duración del contrato, tipo de soporte a la solución y los volúmenes de transacciones esperados. Algunos proveedores ofrecen distintos escenarios de costos dependiendo de si se trata de una PoC o trial o directamente un escenario de producción (tabla 5-3).

Tabla 5-3. Escenarios de costos recurrentes mensuales de psicometría

Volumen de consultas o transacciones mensuales	Trial, prueba de concepto o piloto	Producción (USD)
1-3.000	Menos de 1 mes: generalmente incluidas Más de 1 mes: entre 500 y 2.000 USD	Entre 3.000 y 5.000
3.001 – 12.000	Entre 1.000 y 3.000 USD	Entre 4.000 y 6.000
12.001 – 30.000	Entre 2.000 y 4.000 USD	Entre 5.000 y 7.000
30.001 – 50.000	N/A	Entre 6.000 y 8.000
50.001+	N/A	Entre 9.000 y 15.000

Fuente: elaboración propia, basada en investigación previa, bases de datos y estudios de cliente oculto.

Los servicios que generalmente están contemplados en el costo recurrente incluyen el volumen de resultados contratados, alcance de resultados contratados (puntajes, *insights*, variables, también llamada “variable estadística”, etcétera) y soporte estándar a la solución de acuerdo con los ANS.

Algunos proveedores ofrecen la opción de costo por consulta o la capacidad de pagar consultas adicionales, en el evento que excedan el volumen que se tiene contratado en el plan. Existe mucha variabilidad en este costo unitario por consulta como se puede ver en la tabla 5-4.

Tabla 5-4. Escenarios de costos recurrentes por consulta de psicometría

Volumen de consultas o transacciones mensuales	Precio por consulta o transacción (límite inferior en USD)	Precio por consulta o transacción (límite superior en USD)
1-3.000	1,33	2,50
3.001 - 12.000	1,15	2,00
12.001 - 30.000	0,42	0,58
30.001 - 50.000	0,20	0,27
50.001+	0,09	0,18

Fuente: elaboración propia, basada en investigación previa, bases de datos y estudios de cliente oculto.

c. Servicios adicionales

Algunos proveedores ofrecen opciones de consultoría o asesoría en aspectos técnico-analíticos que pueden resultar de utilidad para maximizar el éxito en una implementación. Si bien la mayoría de estos servicios son cotizados bajo un esquema de tiempo y materiales¹⁷, se identificaron algunos servicios que son los más frecuentes entre los proveedores, algunos de los cuales pueden ser pagados de manera única, según ejecución o en pagos diferidos generalmente dentro del costo de licenciamiento mensual (tabla 5-5).

¹⁷ Un contrato por tiempo y materiales se caracteriza porque el cliente solo paga por el tiempo y los materiales efectivamente empleados por el proveedor en el desarrollo del proyecto.

Tabla 5-5. Servicios más frecuentes entre proveedores, pagados de manera única

Servicio	Pago único	Pago diferido mensual
Construcción de un nuevo modelo* psicométrico	Entre 20.000 y 35.000 USD	Entre 2.000 y 4.000 USD
Revisión de desempeño de un modelo	Entre 1.000 y 5.000 USD	Entre 300 y 1.500 USD

* Por lo general cubre el entrenamiento de un nuevo modelo para un producto o segmento específico del cliente.

Fuente:

6. Datos de open banking

6.1. Generalidades en Colombia y en el mundo

6.1.1. Definición

Open banking está transformando la industria de los servicios financieros. Al romper los silos de datos de los bancos tradicionales, *open banking* posibilita que las instituciones financieras, los *fintech* y otros innovadores accedan a los datos de los clientes, incluidos los de sus transacciones bancarias y no bancarias, y utilizar estos datos para desarrollar nuevos productos y servicios que se adapten mejor a las necesidades de los consumidores (Plaitakis y Staschen, 2020).

6.1.2. En LATAM y el mundo

En todo el mundo, esta tendencia está evolucionando de diferentes maneras: (a) En la Unión Europea (UE), el Reino Unido, Corea del Sur, Australia e India, los reguladores tomaron la iniciativa y han ordenado a los bancos abrir sus cuentas de clientes a otras empresas para estimular la competencia. (b) En Estados Unidos y China, *open banking* es un movimiento dirigido por el mercado mediante empresas que establecen relaciones bancarias abiertas entre ellas. (c) Singapur está utilizando una combinación de ambos (figura 6-1).

Figura 6-1. Nivel de adopción de open banking en países con modelos liderados por el regulador^{18, 19}



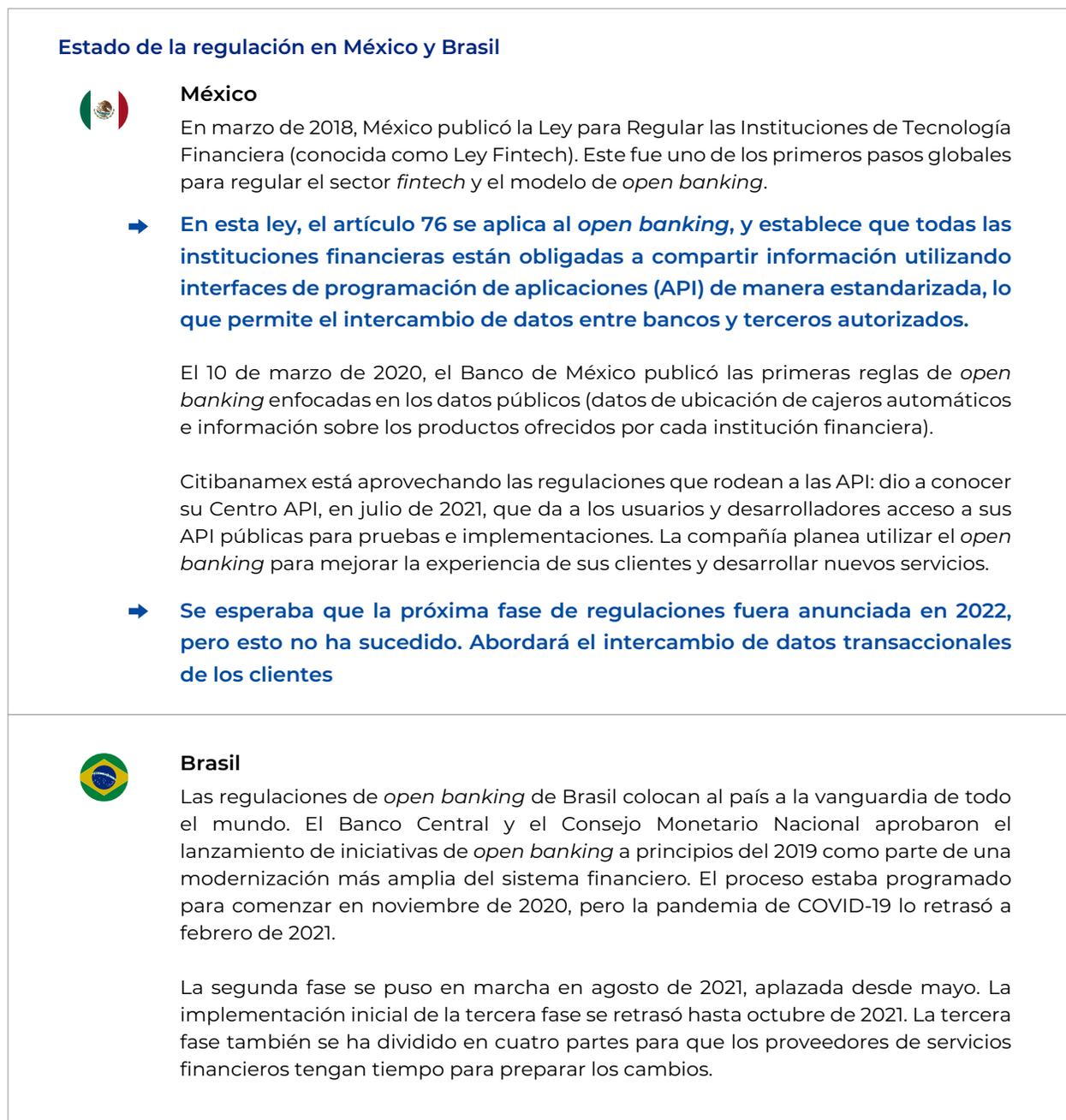
Fuente: tomado y adaptado de (Hong Kong Applied Science and Technology Research Institute, 2020)

¹⁸ La UE y el Reino Unido adoptaron la banca abierta en enero de 2018, Corea del Sur presentó la banca abierta en diciembre de 2019 y Australia, en julio de 2020.

¹⁹ Singapur presentó SGFinDex en diciembre de 2020, una infraestructura público-privada común para permitir datos *open banking* (con datos agregados) para consumidores. Si bien no existe un esquema oficial de licencias/regulación de banca abierta, Singapur publicó un manual de las API para alentar a los bancos a abrir sus datos, sistemas y servicios, y ha otorgado licencias de banca digital a una variedad de terceros, por ejemplo, Grab (Uber), Sea.

Mientras tanto en Latinoamérica, el *open banking* ha avanzado rápidamente, aunque de manera desigual: Brasil, se encuentra en fase de implementación; mientras que, en México, la Ley Fintech de 2018 contempla el *open banking* como un artículo clave en la regulación del ecosistema (Belvo, 2022) (tabla 6-1 y figura 6-2).

Figura 6-2. Estado de la regulación en países selectos de LATAM



Fuente:

En Perú, el *open banking* está en una etapa temprana, pero en los próximos años podría convertirse en un escenario oportuno para que entidades financieras trabajen de la mano en beneficio de los consumidores peruanos. Aunque el Banco Central no ha establecido ninguna regulación específica para *open banking*, las instituciones financieras ya están compartiendo datos de los clientes. Cada mes, las instituciones financieras peruanas envían información del cliente al Banco Central para crear un perfil. Esto podría considerarse una ventaja inicial ventajosa si se compara con otros países latinoamericanos. Sin embargo, falta una consideración muy importante en esta situación: el consentimiento del cliente que permitiría a ciertas IF acceder a la información (BBVA, 2022).

Figura 6-3. Cuatro fases de la implementación de la regulación de Brasil



Fuente: elaboración propia con información publicada por Belvo y gobierno de Brasil.

En Colombia la regulación del *open banking* aún se encuentra en una fase temprana, aunque ya se ha emitido y expedido un decreto que ampara el desarrollo de estos modelos bajo las actuales leyes de protección de datos.

Sin embargo, a finales de 2021 se publicaron una serie de documentos que sentaron las bases preliminares para el funcionamiento de este modelo en Colombia. Adicional a esto, en julio de 2022, se expidió el Decreto 1297 (25 de julio de 2022) (Ministerio de Hacienda y Crédito Público, 2022), que busca regular la transferencia de datos del consumidor entre entidades financieras.

6.2. Fundamento teórico

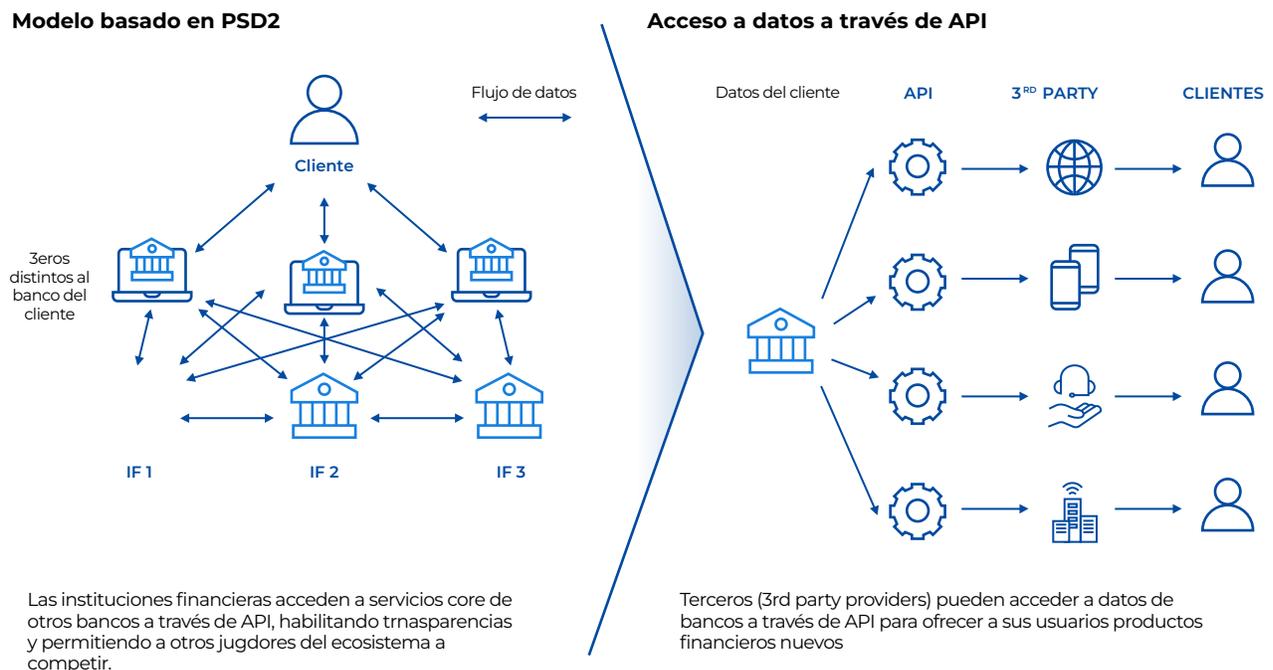
Provocados por una combinación de regulación gubernamental y fuerzas del mercado, los datos financieros abiertos posibilitan que un universo en expansión de jugadores, tanto financieros como no financieros, acceda a las cuentas y los datos de los clientes para ofrecer nuevos productos y servicios (McKinsey y Company, 2021).

El *open banking* surge en 2016, cuando los reguladores de la Unión Europea presionaron por primera vez para crear datos financieros abiertos, sentando las bases para el desarrollo del mercado. Esta pieza radical de la legislación de la UE (conocida como la segunda Directiva de Servicios de Pago, o PSD2) también impulsó a la Autoridad de Mercados y Competencia (CMA) del Reino Unido a ordenar el desarrollo de un “estándar de banca abierta” por parte de los nueve bancos más grandes del país para compartir información de los clientes y datos de transacciones con terceros.

Fundamentalmente, el *open banking* depende de una tecnología llamada “protocolo API”. Esta tecnología habilita el intercambio de información de manera estándar entre participantes de un ecosistema (en el caso colombiano, las entidades reguladas) y otros proveedores de servicios (IMF y Fintech, y otros).

Los terceros (figura 6-3), tienen acceso aprobado a las API que habilitan los miembros del ecosistema, incluyendo las interfaces para obtener y procesar los datos de los usuarios de servicios financieros (Qualco, 2022).

Figura 6-4. Modelo de compartición de datos basado en PSD2*



* PSD2 = Segunda directiva de servicios de pago en la Unión Europea.
Fuente: McKinsey y Company, 2021.

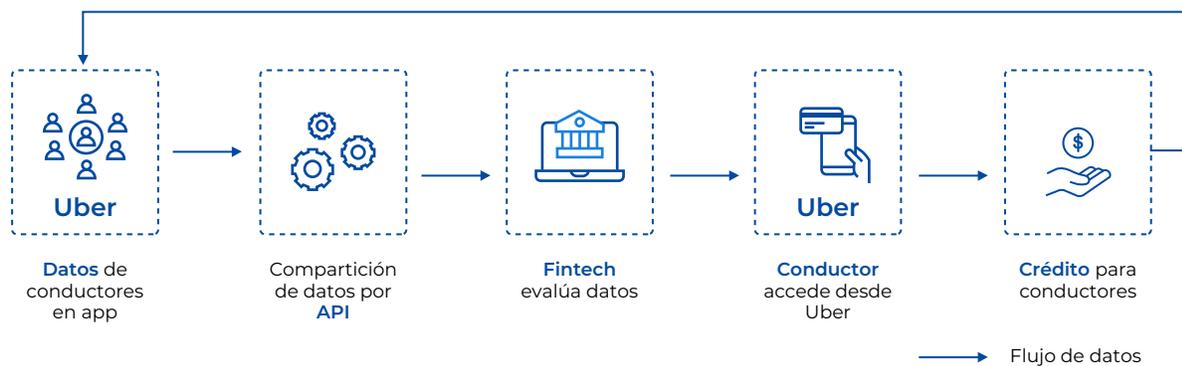
6.3. Modelos de negocio posibles a través del open banking

El *open banking* abre nuevas posibilidades a las instituciones financieras para el desarrollo de nuevos modelos de negocio (SBS, 2022). Un banco tradicional ofrece sus productos y los distribuye a través de sus canales: sucursal, sitio web o aplicación móvil. Sin embargo, existen modelos en los que quien desarrolla el producto y quien lo distribuye son distintos actores.

Modelo 1. Finanzas embebidas

A través de un agregador de datos de *open banking* conectado mediante las API a sistemas bancarios y otros participantes, un tercero puede acceder a información de clientes para brindar servicios financieros. Por ejemplo, en México, una IF ofrece cuentas y tarjetas de crédito a conductores de Uber a través de la *app* de Uber, que comparte datos de identificación e ingresos mediante las API para facilitar la evaluación crediticia. En este caso, el producto financiero se ofrece a través de un tercero utilizando la *app* de Uber como canal (figura 6-4).

Figura 6-5. Modelo 1. Embedded Finance (finanzas embebidas)

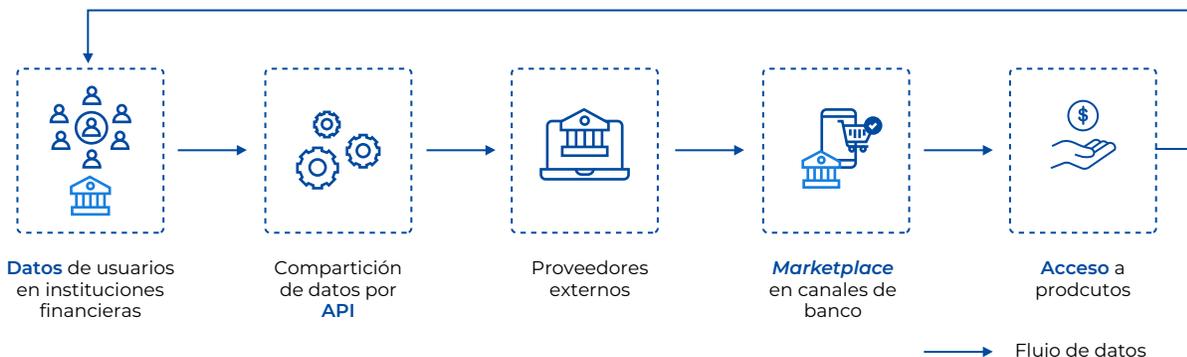


Fuente: elaboración propia.

Modelo 2. Marketplace

En contraposición a finanzas embebidas, en este enfoque, un proveedor de servicios financieros aprovecha sus canales para promocionar productos de terceros. Por ejemplo, un banco incorpora en su aplicación móvil la posibilidad de comprar seguros o invertir, así como productos no financieros, por ejemplo, viajes. Estos terceros crean productos utilizando los datos de la entidad financiera, proporcionados a través de API (figura 6-5).

Figura 6-6. Productos de terceros en marketplace de banco



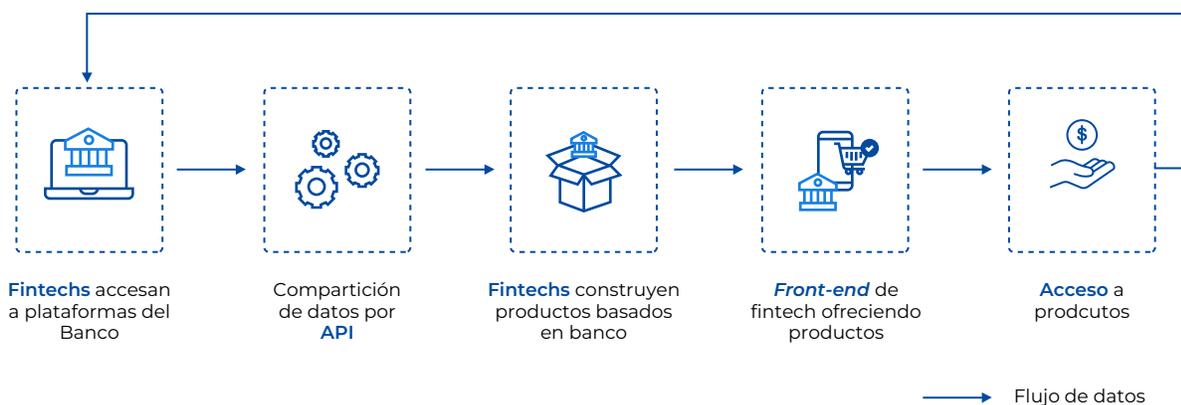
Fuente: elaboración propia.

Modelo 3. Banking as a Service (BaaS) (banca como servicio)

El banco no desarrolla el producto ni lo distribuye, sino que ofrece sus capacidades a otra empresa para que los explote. Así, terceros pueden ofrecer servicios financieros gracias a la infraestructura, tecnología y licencia bancaria provista por el banco. Por ejemplo, el banco puede proveer el servicio de *onboarding* (KYC), *credit scoring*, pagos a través de QR, tarjetas de débito y crédito, entre otros. En algunos países de la región, se observa que los bancos ofrecen estos servicios a través de sus API *marketplace* (portal de desarrolladores) (Business Insider, 2020).

Los *fintech* (glosario de términos) cumplen con un rol importante en estos nuevos modelos de negocio. Estas pueden ser habilitadores, proveyendo servicios especializados de asesoría a los bancos u otras entidades financieras que quieren abrirse y requieren apoyo en el desarrollo de las API. Los *fintech* son también consumidores de las API, lo que les posibilita ofrecer servicios de valor añadido (figura 6-6). Ejemplo: agregadores de cuentas. El M-POS Clip en México usa las capacidades de cuenta de un banco grande para ejecutar la adquirencia (guardar el dinero cobrado), mientras que Clip se centra en desarrollar los datáfonos móviles.

Figura 6-7. Banking as a service, modelo



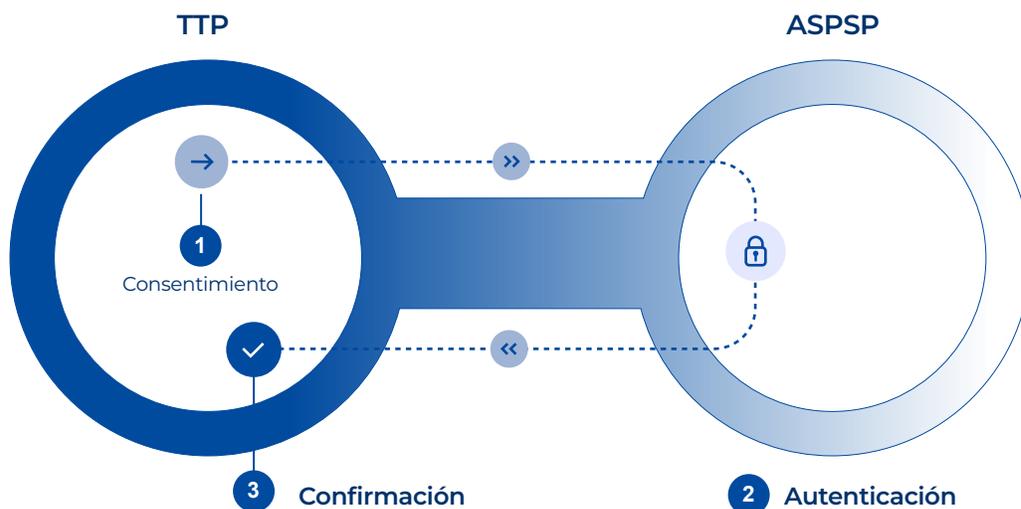
Fuente: elaboración propia.

6.4. Proceso de recolección de los datos

Como se muestra en la figura 6-7, los datos de *open banking* que tienen las IF y que comparten a terceros por medio de las API. Los terceros pueden conectarse a dichas API directamente o a través de agregadores. Una vez que se tienen los datos, los terceros podrán ser capaces de manipularlos para producir nuevos productos financieros, implementar estrategias y hacer *scoring* alternativo, etcétera.

Por otro lado, el consentimiento del consumidor a lo largo del recorrido de *open banking* es vital. Está construido por diseño a través de la autenticación, los datos transaccionales y los procesos de entrega de pagos (figura 6-7).

Figura 6-8. Flujo de autenticación simple, basado en el OBIE del RU*



*OBIE = Open banking. TPP= 3rd party provider (proveedores de servicios) y ASPSP = Account Servicing Payment Service Providers (quienes proveen y mantienen los datos de los usuarios).

Fuente: (Qualco, 2022)

Los datos de *open banking* pueden ser obtenidos relativamente en corto tiempo. De acuerdo con Belvo en un máximo de 10 segundos se pueden procesar gran cantidad de transacciones de manera que las IF reciben datos limpios, categorizados (p. ej. categoría de gasto) y detallados (p. ej. nombre del comercio) listos para incorporar en los modelos (Belvo, 2023).

6.5. Disponibilidad, cobertura y veracidad de datos

Disponibilidad: como se indicó anteriormente los datos de *open banking* requieren del otorgamiento de permisos que proporcionen directamente los usuarios, ante lo cual pueden existir barreras de acceso debido a la desconfianza de los usuarios para compartir sus credenciales. Por otro lado, si el proveedor o agregador de datos tiene configurados los servicios de las cuentas bancarias que en efecto el usuario posee se pueden obtener datos de las últimas transacciones realizadas por el usuario y en algunos casos hasta de 24 meses. Por lo que si el usuario tiene dichas cuentas se pueden obtener la cantidad y la vigencia adecuada de datos.

Cobertura: la cobertura de los datos de *open banking* depende si el proveedor o agregador tiene numerosos convenios de agregación de datos con las principales instituciones financieras dentro de la población objetivo en el segmento deseado, luego entonces dicho proveedor tendrá buena cobertura.

De acuerdo con el estudio de “Demanda de productos financieros en los micronegocios de 2019”, el 20,1 % de los microempresarios ahorran en una institución financiera (Banca de las Oportunidades, 2020).

Sin embargo, vale la pena también considerar tendencias más recientes como las de las billeteras digitales que han tenido un crecimiento sostenido muy interesante; entre el año 2020 y 2021 fue de 195 %, estimaciones recientes indican que Nequi y Daviplata tienen más de 20 millones y 22 millones de usuarios, respectivamente²⁰ la penetración de billeteras digitales, incluyendo la billetera Movii, asciende al 76 % de la población (Mora, 2023), ya que están en el día tras día de las finanzas personales de una parte importante de la población, desde pagos de transporte público, la aceptación de códigos QR, los pagos de beneficios sociales, las compras en el mercado en la aplicación, el pago de facturas y recargas móviles, y las remesas internacionales hasta la compra de criptomonedas.

Con base en lo anterior, aún sin conocer la proporción exacta de los micronegocios que usan billeteras como medio de pago, se puede suponer que los proveedores de *open banking*, que tengan convenios activos con dichas billeteras podrán mejorar sustancialmente la abertura de los datos de *open banking* para el segmento de las EPC. Es frecuente que estas billeteras sean también usadas como medio de pago para compras de bajo y mediano valor, principalmente en micronegocios. Por ejemplo, DaviPlata creó la herramienta “Perfil Negocio”, en la cual a través de catálogos digitales se puede dar a conocer su oferta de productos completa y ofrecer otras alternativas de pago, como links de PSE y códigos QR (Colombia Fintech, 2022), más recientemente.

Especificidad: las fuentes de *open banking* pueden proporcionar información específica sobre el individuo a ser evaluado, sin embargo pueden contener algunos sesgos ya que se espera que para el segmento de las EPC no haya segregación entre las transacciones específicas de su actividad económica y las del emprendedor como persona; pese a ello, las capacidades de las fuentes de *open banking* de lograr resultados enriquecidos como estimadores de ingresos, simuladores de capacidad de pago por usuario lleva a que tengan muy buena especificidad para casos de uso de otorgamiento de crédito.

6.6. Resumen de beneficios y retos potenciales

Las IF, centrales de información, agencias/firmas de analítica de datos e inclusive la banca pública de desarrollo, podrán obtener diversos datos a través de un agregador con las API de los bancos aliados. Así, podrán potenciar su negocio validando la identidad de una persona con toda la información que ya conocen los bancos minimizando su tasa de fraude, así como evaluar la viabilidad del crédito.

Un estudio realizado por la consultora McKinsey, en el que se comparó el desempeño de los modelos de riesgo tradicionales con los que usan datos de *open banking* en varios mercados europeos, se encontró que el uso de datos de *open banking* mejoró la precisión de las predicciones de riesgo en un 10 vs. en promedio (Birch, Cummins, y Shin, 2018).

²⁰ Con datos recabados de Data.ai del 01/03/2023, y contando 1,2 descargas por usuario (Data.ai es un rastreador de aplicaciones que, entre otros datos, mide el número de descargas y usuarios activos de aplicaciones móviles).

De la misma manera, relevantes centrales de información como Equifax y Experian también han identificado el uso de datos de *open banking*, puesto que proporcionan una visión más completa y precisa sobre la situación financiera de los consumidores y las empresas que pueden mejorar la precisión de los modelos de *scoring* de crédito para préstamos a pequeñas empresas (Experian, 2018). Los resultados muestran que el uso de datos de *open banking* puede aumentar la precisión del puntaje alrededor del 20 % (Equifax, 2019).

Existen también beneficios específicos que deberían propiciar un intercambio de información de este tipo, los cuales están resumidos en la tabla 8-2.

Tabla 8-2. Beneficios y retos percibidos

Caso de uso	Beneficios	Retos
Automatización completa de ingresos y gastos	<ul style="list-style-type: none"> • Reduce costos de análisis • Mejora precisión del análisis 	<ul style="list-style-type: none"> • Tasa de adopción • Categorización de gastos • Conciliación del flujo de caja • El gasto histórico no es igual al presupuesto futuro
“Prerellenar” ingresos y gastos en solicitud	<ul style="list-style-type: none"> • Reduce costos de análisis • Mejora precisión del análisis • Satisfacción del cliente 	<ul style="list-style-type: none"> • Acción humana, implica costos que contrarestren los ahorros generados en otros puntos • Integración tecnológica
Validación de ingresos	<ul style="list-style-type: none"> • Mejora del recorrido del cliente • Mayor precisión de ingresos • Velocidad de validación (vs.papel) 	<ul style="list-style-type: none"> • Múltiples fuentes de ingresos no reflejadas en un solo estado de cuenta • Integración de procesos
Análisis de gastos fijos	<ul style="list-style-type: none"> • Mejora de la precisión del presupuesto • Buen grado de especificidad entre el micronegocio y el emprendedor 	<ul style="list-style-type: none"> • Caso de uso limitado, solo un elemento de ingresos y gastos, aún se requieren detalles adicionales
Análisis de variables	<ul style="list-style-type: none"> • Mejora de la precisión del presupuesto 	<ul style="list-style-type: none"> • La categorización y precisión de los gastos es difícil
Predicción del flujo de caja	<ul style="list-style-type: none"> • Alerta temprana en tiempo real de problemas de flujo de efectivo • Mejores resultados del plan de pago • Buen complemento para análisis de capacidad de pago 	<ul style="list-style-type: none"> • Puede que no refleje todos los ingresos o gastos
Detección de vulnerabilidades	<ul style="list-style-type: none"> • Identificación proactiva de posibles vulnerabilidades 	<ul style="list-style-type: none"> • Requisitos de acción operativa (qué hacer con la información sin consentimiento) • Gobernanza de datos sensibles y su impacto en el cumplimiento de protección de datos
Revisión de desencadenantes	<ul style="list-style-type: none"> • Disparadores basados en información de gastos en tiempo real • Acción proactiva en caso de que cambie la situación financiera de los clientes 	<ul style="list-style-type: none"> • Límite de autenticación de 90 días • La sensibilidad de los datos puede socavar la confianza del cliente

Fuente: elaboración propia, con base en Qualco (2022) y McKinsey y Company (2021).

6.7. Principales proveedores

Similar a la fuente de datos de psicometría, para *open banking* se realiza una preidentificación de posibles soluciones a partir de su oferta de valor. Muchas ofertas dirigidas por bancos en este momento se centran en gran medida en la agregación de diferentes cuentas en un solo lugar, y las entidades no bancarias a menudo brindan superposiciones funcionales además de la agregación. Dichas propuestas se dividen en dos categorías principales: proveedores de infraestructura y propuestas orientadas al cliente, y las propuestas orientadas al cliente se dividen aún más en aumentos de productos existentes y experiencias de cliente completamente nuevas (figura 6-8 y tabla 6-3).²¹

Figura 6-9 Tres tipos de proveedores de open banking



Fuente: elaboración propia con análisis e identificación.

²¹ Esta sección está basada en estudios desarrollados en Gran Bretaña, debido a que es el país más avanzado. Soluciones presentes en LATAM y, particularmente en Colombia, siguen modelos y propuestas de valor similares.

Tabla 6-3. Preidentificación de proveedores open banking

Compañía	HQ	Categoría	Sitio web	Año de fundación	Experiencia Colombia	Comentarios
Akropoli ★	Brasil	Open banking	akropoli.tech	2020	No	Enfoque en crédito, presencia en Colombia.
Finerio Connect ★	México	Open banking	finerioconnect.com	2016	Si	Solución robusta, con red de <i>partners</i> en Colombia.
BitCapital	Brasil	Open banking	bitcapital.com.br	2018	No	N/A
MO Tecnologías	Colombia	Open banking	wearemo.com	2016	Sí	Plataforma basada en API enfocadas en incrementar acceso a crédito.
Cirenio ★	Argentina	Open banking	cirenio.com	2021	Sí	Solución robusta y con enfoque en crédito.
Belvo ★	México	Open banking	belvo.com	2019	Sí	Plataforma API de finanzas abiertas.
Knowtech	Argentina	Open banking	knowtech.com.br	2001	Sí	Enfoque en desarrollo de productos, tienen varias licencias en distintos países.
Live On Solutions	Brasil	Open banking	liveonsolutions.com	2012	No	No relevante
Poincenot ★	Argentina	Open banking	poincenot.com	2014	Sí	Solución de crédito enfocada en persona y negocio. Ofrecen una solución integrable.
Teddy Open Finance	Brasil	Open banking	teddydigital.io	2020	No	No relevante.
Teros	Brasil	Open banking	teros.com.br	2014	TBD	Solución robusta para crédito, por confirmar presencia en Colombia.
Yunius	México	Open banking	yunius.com	1999	No	Principalmente Core Banking System.
Guiabolso	Brasil	Open banking	guiabolso.com.br	2012	No	Solo Brasil.
Celcoin	Brasil	Open banking	celcoin.com.br	2015	No	Solo Brasil.
Quanto	Brasil	Open banking	quan.to	2016	No	Player interesante por <i>funding/investors</i> y propuesta, en expansión en LATAM.
Franq	Brasil	Open banking	franq.com.br	2019	No	No relevante.
Pluggy	Brasil	Open banking	pluggy.ai	2019	No	Propuesta robusta para crédito.
CashWay	Brasil	Open banking	cashway.io	2019		No relevante.

Tabla 6-3. Preidentificación de proveedores open banking (continuación)

Compañía	HQ	Categoría	Sitio web	Año de fundación	Experiencia Colombia	Comentarios
Transfeera	Brasil	Open banking	transfeera.com	2017	Sí	No relevante.
Aarin (Acquired by Next)	Brasil	Open banking	aarin.com.br	2020	No	Pix.
EOS Software	México	Open banking	eossoftware.mx	2018	TBD	Presencia en México. Enfoque en crédito, por confirmar actividad en Colombia.
Wall Street	México	Open banking	portalwallstreet.com	2017	No	No relevante.
Datanomik	Uruguay	Open banking	www.datanomik.com	2021	Sí	
CIAL ★	México	Open banking	www.cialdnb.com	2016	Sí	Empresa en expansión agresiva en LATAM. Presencia en Colombia, soluciones de crédito para SME.

★ Proveedores preseleccionados.

Fuente: elaboración propia con base en investigación previa, bases de datos y asociaciones Fintech.

6.8. Análisis de propuestas comerciales

Se identificaron dos proveedores que tienen operaciones en Colombia: Belvo y Finerio Connect. Derivado de las conversaciones que se sostuvieron, es posible presentar un aproximado de costos de integración y operación de una plataforma de *open banking*. Este estimado se entrega con el fin de comprender las estructuras de costos que los proveedores ofrecen típicamente y que el piloto deberá asumir en el cálculo de costos y esfuerzos requeridos (resumido en tabla 6-4).

6.8.1. Inversión estimada

De manera similar a datos de psicometría, (sección 5.9), la inversión en soluciones de *open banking* está asociada a las etapas del proyecto de implementación y requerimientos mencionados anteriormente. También se identificó que existen tres momentos de inversión: instalación inicial, costo recurrente y servicios adicionales. A continuación, se describe cada uno de ellos con unos rangos de costo.

a. Instalación inicial

Los proveedores consultados establecen una cuota por integración de las API que dependen del nivel de complejidad de instalación. Como se describe en la tabla 6-4, el precio varía entre US 1.000 (~4,8 millones COP) a US 2.000 (~10 millones COP).

Tabla 6-4. Escenarios de costos de instalación open banking

Concepto	Complejidad baja	Complejidad alta
Costo por integración a API de agregación (completa)	1.000 USD	2.000 USD

Fuente:

b. Costos recurrentes

Los proveedores de *open banking*, tienen cuota mensual que provee acceso a la plataforma, al servicio al cliente, y otros costos de operación que varían entre 500 USD (~2,5 millones COP) a 1.000 USD (~4,8 millones COP) (tabla 6-5). El precio se establece con relación al tamaño de la organización contratante y del volumen de uso esperado.

Tabla 6-5. Escenarios de costos recurrentes mensuales de open banking

Concepto	Mínimo	Máximo	Recurrencia
Suscripción (servicio al cliente, gestión de cuenta, ingeniería)	500 USD	1.000 USD	Mensual

Fuente:

c. Costos por consulta

Finalmente, se cobra una cuota fija por cada consulta realizada, que puede ir desde 0,1 USD (480 COP) a 0,7 USD (3.400 COP), que incluye la revisión completa de los puntos de datos bancarios, fiscales, enriquecimiento de la consulta y categorización (evaluación).

En total, el costo estimado anual, considerando mil consultas mensuales es de 15.400 USD (~7,4 millones COP), o considerando 10.000 mil consultas mensuales es de 95.000 USD (~457 millones COP) (tabla 6-6).

Tabla 6-6. Escenarios de costos por consulta de open banking

Concepto	Mínimo	Máximo	Recurrencia
Agregación bancaria	-	0,20 USD	Por revisión
Enriquecimiento	-	0,30 USD	Por revisión
Categorización	-	0,10 USD	Por revisión

Fuente: estimaciones basadas en ofertas comerciales vigentes de proveedores de *open banking* a febrero de 2023.

6.9. Conclusiones *open banking*

Open banking ofrece a las instituciones financieras colombianas un medio para expandir la información transaccional de la cual se dispone a la hora de evaluar el crédito. Con esta nueva información, es posible llegar a conclusiones respecto al negocio que brinda un mayor entendimiento sobre la situación real del solicitante y tomar una decisión crediticia con menor probabilidad de impago y con mayor velocidad.

En Colombia existe un número de agregadores que ya tienen integraciones y accesos a las API y datos de entidades como los bancos, las SEDPE, las IF, billeteras digitales, Fintech y otros, que además ofrecen mecanismos de evaluación embebidos, así como integraciones directas (o incluso acceso mediante *links* que eliminan barreras tecnológicas) y estándares de seguridad alineados con ISO 27001.

Esto permitirá a las IF, banca de desarrollo o agencias/firmas de analítica de datos establecer integraciones relativamente sencillas con agregadores para determinar viabilidad del crédito de algunos solicitantes. Sin embargo, algunos retos se mantienen: costos, personal y capacidades técnicas y disponibilidad de datos.

7. Datos de telefonía móvil (Telco/CDR)

7.1. Generalidades

A través del registro del número telefónico del cliente, es posible conocer y evaluar el comportamiento que el solicitante ha tenido con su línea telefónica. Algunos de los datos a los que se pueden acceder son:

- **Prepago/pospago:** información de activación de la tarjeta SIM o tenencia del cliente junto con otros detalles mencionados en el formulario de solicitud o contrato. Información de morosidad, crédito, CHURN (tasa de abandono) y pagos de pospago. Disponibilidad de billeteras móviles.
- **Uso de datos:** datos utilizados, ingresos generados a partir de los datos, uso de datos por hora, servicios de valor añadido relacionados con los datos, detalles adicionales sobre aplicaciones utilizadas, sitios web navegados durante el día y la noche.
- **Geolocalización:** movilidad a partir de datos CDR (Call Detail Record, por sus siglas en inglés) (registros detallados de llamadas), presencia diurna y nocturna, densidad de ubicación, ubicación destacada con atributos del censo, datos públicos disponibles.
- **Historial de recarga/top ups:** información de recarga, tipo, tamaño y frecuencia de recarga, canal de pago de facturas, monto facturado, condiciones de pago y modo de pago, información bancaria, de crédito y de monedero.

- **Patrones de llamadas y SMS:** estadísticas sobre duración/recuento de llamadas, números de teléfono que llaman, torres, SMS (short message service, por sus siglas en inglés) (servicio de mensajes cortos o servicio de mensajes simples) enviados y recibidos, hora del día de las llamadas, día de la semana/fin de semana e inactividad.

7.2. Experiencias internacionales

En un piloto realizado por Claro Brasil, se usaron técnicas de ML y a través de datos de clientes prepago, como recargas, historial, uso de datos, uso de voz, regiones, entre otras se obtuvo un puntaje de crédito de alta precisión sin dependencia del puntaje del buró y una precalificación de los clientes (Dietrich, De Souza, y Guerreiro, 2020).

El potencial de resultados cuando se aplican datos de telecomunicaciones en la calificación crediticia es muy alto, pero para capturar esta oportunidad hay desafíos técnicos de ciencia de datos, desafíos legales por tema de cumplimiento normativo y privacidad además que se requiere de un MNO (Mobile Network Operator, por sus siglas en inglés) (operador de red móvil) que suministre la información, esto hace que la disponibilidad y accesibilidad a estas fuentes de información decrezcan dramáticamente.

En el caso mencionado en Brasil, la morosidad del primer pago se redujo en un 31 % y las ventas aumentaron 11 %. Más allá de los resultados internos, este modelo predictivo combinado con otros modelos de comportamientos permitió presentar al mercado el Claro Score.

7.3. Resumen de beneficios potenciales

Esta fuente de datos ofrece una cobertura amplia, ya que más del 90 % de las personas tiene línea telefónica, que en teoría puede ser posible analizar, de esta manera, permite tres beneficios principales:

- **Análisis de riesgo asertivo:** el análisis de los datos de telefonía móvil dibuja una imagen acertada sobre el comportamiento del solicitante. Es decir, predice con cierto rango alto de certeza, la probabilidad de impago del solicitante.
- **Estrategias de cobranza:** las instituciones de crédito pueden aplicar análisis de datos a los datos de telefonía de los clientes para comprender mejor su comportamiento y características para maximizar los rendimientos de cobro. En el proceso de cobranza tradicional, las IF segregan a los clientes en algunas categorías basadas en riesgo y establecen estrategias de contacto para cada una. Mediante el uso de datos de telefonía móvil, las IF pueden avanzar hacia una comprensión más profunda y matizada de sus clientes. Con base en fuentes de datos que incluyen datos demográficos del deudor, actividad de la línea, cobranzas y calificaciones de riesgo, los clientes se pueden clasificar en micro segmentos y pueden llevar a cabo intervenciones más efectivas.

- Detección de fraude: el análisis de estos datos se ha convertido en una herramienta importante para identificar solicitudes fraudulentas, ya que puede ayudar a validar la identidad y la autenticidad de un cliente. El análisis de datos de aplicaciones móviles ofrece una verificación continua de posibles escenarios de fraude incluso después de que se haya aprobado el préstamo.

7.4. Resumen de retos potenciales

A pesar de las ventajas y la capacidad de elaborar modelos con un rango de certeza amplio, existen distintos riesgos que mitigan los beneficios. Entre ellos:

- La disponibilidad de los datos se puede ver reducida debido a regulaciones locales de protección de datos y privacidad, ya que pueden disminuir o restringir por completo el acceso a datos crudos y procesados de valor.
- La disponibilidad está disminuyendo. A pesar de que se tiene una gran cobertura de líneas telefónicas, los patrones de uso o consumo cambian constantemente. Cada vez se usan más los servicios VoIP (Voice Over Internet Protocol, por sus siglas en inglés) (voz sobre protocolo de Internet) para realizar llamadas por ejemplo Whatsapp Call o Whatsapp Video Call. Que sustituyen a los datos de Telco/CDR.
- Se ha identificado que establecer convenios interinstitucionales entre los proveedores de telefonía, agregadores, agencias/firmas de analítica de datos y FI puede resultar desgastante en términos jurídicos y comerciales, además que supone inversiones relevantes por adelantado. En particular, en Latinoamérica, los operadores móviles son instituciones grandes, con penetración importante en el mercado y que usualmente son muy reservados con su información.
- El punto anterior impacta también en la cobertura: para tener la cobertura deseada se deberían implementar convenios con los diferentes proveedores de telefonía o identificar agregadores que ya cuenten con las relaciones.
- Especificidad limitada en el caso de mipymes: cuando se trata de mipymes y EPC constantemente se tiene la duda de si la actividad y relacionamiento que se desea medir es la del dueño de la EPC, como persona o como parte de su actividad económica.
- Una misma EPC puede tener una o varias cuentas o líneas tanto de atención como de operación.

En algunos mercados los operadores de telefonía han realizado alianzas con entidades financieras con el fin de desarrollar productos de crédito, lo cual supone exclusividad de ciertos datos hacia dichas entidades. Otros han incursionado al mundo de préstamos debido a su amplia base de clientes, cantidad de información recolectada, conocimiento de clientes, y han también ampliado capacidades de cobro que han podido ser fácilmente trasladadas a las prácticas de cobranza de créditos.

El principal enfoque de estas alianzas ha sido el crédito de consumo. Sin embargo, las EPC han sido incluidas, al pertenecer a un segmento intermedio entre persona natural y mipymes. Tal es el caso de la tarjeta de crédito Claro Pay Mastercard.²²

Otro caso es el de Movistar Money, que tiene como socios a Telefónica y BBVA, donde cliente pueden acceder a créditos tanto de equipos como de libre inversión hasta por 12 SMMLV, lo cual puede ser un monto atractivo para una EPC de operación principalmente urbana si es otorgado al dueño de este micronegocio. Al ser de libre inversión muy seguramente será catalogado como crédito de consumo, pero su uso y destinación real serán muy similares a la de microcrédito.

8. Datos de dispositivos *smartphone* (extraídos por SDK o similar)

8.1. Generalidades

Se refiere a los datos rastreables de un dispositivo móvil, a los cuales se accede a través de una aplicación después de que el usuario haya otorgado permiso. Es decir, se requiere que una aplicación móvil esté instalada en el dispositivo para que los datos puedan ser recolectados.

Colombia presenta una penetración de *apps* de servicios financieros y bancarios del 46,5 %²³. Esto denota que las personas están adquiriendo habilidades de uso de servicios financieros móviles. Sin embargo, solo el 4,7 % de los usuarios mayores de 15 años tienen una cuenta bancaria móvil²⁴. El constante uso de las *apps* en los dispositivos móviles lleva a que la cobertura de las poblaciones crezca. De la misma manera, el uso de las súper *apps* ha cambiado la forma en que se piensa sobre las interacciones entre los usuarios y el comercio. Entonces no sorprende que también esté redefiniendo la forma en que se hace la banca (Roa, Correa-Bahnsen, y Rodríguez-Rey, 2021).

En un estudio que se realizó en Colombia se investigaron diferentes interacciones entre usuarios dentro de Rappi²⁵ para proporcionar un nuevo conjunto de datos que pudiera ser usado para predecir el comportamiento del prestatario. Se analizaron cinco tipos de interacciones del usuario dentro de la súper app: (a) transferencias P2P (*peer-to-peer*, por sus siglas en inglés) (persona a persona, red de pares, red entre iguales o red entre pares); (b) relaciones de usuarios que usan una misma tarjeta; (c) el mismo dispositivo; (d) compartir la misma geolocalización basada en las direcciones de entrega, y (e) los primeros seis dígitos de una tarjeta (conocido como BIN). El estudio concluyó que las variables (también llamada "variable estadística"), derivadas del BIN (número de identificación bancaria, por sus siglas en inglés) con un efecto alto en la evaluación de crédito; también que la centralidad de tarjetas de crédito y variables de geolocalización se relacionan con el uso de estas tarjetas, uso de descuentos y gasto. De esta manera, se puede concluir que si una IF tiene una *app* lo suficientemente posicionada y con muchas transacciones podrían usar datos del propio comportamiento transaccional de los usuarios para construir variables alternativas que puedan servir para enriquecer sus modelos.

²² Tarjeta de crédito Claro Pay Mastercard, en Colombia, ofrece crédito con ingresos desde un SMLV, múltiples beneficios de compra y consolidación de pago automático Telco+crédito a través de factura.

²³ Fuente: GWI 2020. Las cifras representan los hallazgos de la encuesta general de usuarios de internet entre los 16 y los 64 años. globalwebindex.com

²⁴ Datos de inclusión financiera global del Banco Mundial (WBG) a enero de 2021.

²⁵ Rappi es una plataforma de comercio que opera a través de una *app* móvil y que conecta a los clientes con una red de Rappitenderos. Tiene la capacidad de acercar el consumidor a múltiples bienes o servicios.

Otra manera de recolectar datos de dispositivos móviles es a través de un SDK (*Software Development Kit*, por sus siglas en inglés) (estuche de desarrollo de *software*,) que puede ser de un externo. El mundo de los datos alternativos de dispositivos móviles representa un fragmento de código que se incluye dentro de una *app* móvil y que al ser invocado recolecta datos crudos del dispositivo de acuerdo con los permisos expresos que dé el usuario.

Algunos de los datos que se pueden obtener (Google Play, s.f.). Son:

- **Ubicación:** si el usuario da permisos expresos o también a través de información de ubicación derivada de una dirección IP (Internet Protocol, por su sigla en inglés) (información sobre identificación personal).
 - **Ubicación aproximada:** ubicación física del usuario o dispositivo en un área igual a 3 kilómetros cuadrados o mayor, como la ciudad en la que se encuentra un usuario o la ubicación proporcionada, por ejemplo, el permiso ACCESS_COARSE_LOCATION de Android.
 - **Ubicación precisa:** ubicación física del usuario o dispositivo en un área menor a 3 kilómetros cuadrados, como la ubicación proporcionada por el permiso ACCESS_FINE_LOCATION de Android.
- **Información personal**
 - **Nombre:** corresponde a la forma en que un usuario se refiere a sí mismo, como su nombre, apellido o sobrenombre.
 - **Dirección de correo electrónico:** es la dirección de correo electrónico de un usuario.
 - **ID de usuario:** son los ID de usuario que se relacionan con una persona identificable. Por ejemplo, un ID, número o nombre de cuenta.
 - **Número de teléfono:** el número de teléfono de un usuario. Donde adicionalmente existen indicadores más específicos como por ejemplo registro de llamadas, lectura de números de teléfonos almacenados o estado de teléfono de acuerdo con versión del sistema operativo.
- **Mensajería:** existen diferentes maneras en las cuales la *app*, o una biblioteca incluida en ella, pueden acceder a los datos del usuario relacionados con los mensajes. Algunos ejemplos son:
 - **Correos electrónicos:** incluye los correos electrónicos del usuario, incluidos los asuntos, los remitentes, los destinatarios y el contenido de los mensajes.
 - **SMS o MMS:** incluye los mensajes de texto del usuario, incluidos los remitentes, los destinatarios y el contenido de los mensajes.
 - **Otros mensajes:** otros tipos de mensajes, por ejemplo, mensajes instantáneos o contenidos de chat.
- **Biblioteca multimedia:** existen diferentes maneras en las *apps*, o una biblioteca incluida en ellas, pueden acceder a los datos del usuario relacionados con las fotos, video y archivos de audio entre ellos. Grabaciones de voz o sonido, la voz del usuario o archivos de música.

- **Calendario y contactos:** existen diferentes maneras en las *apps*, o una biblioteca incluida en ellas, pueden acceder a los datos del usuario relacionados con lectura de citas de calendario, creación de citas, lectura de contactos creados, lectura de números telefónicos y creación de contactos.
- **Actividades en apps:** existen diferentes maneras en las *apps*, o una biblioteca incluida en ellas, pueden acceder a los datos del usuario relacionados con interacciones con la *app*, historial de búsquedas, *apps* instaladas y producción de contenido, entre otras.
 - **Interacciones en apps:** información acerca de cómo un usuario interactúa con tu *app*. Por ejemplo, la cantidad de veces que visita una página o qué elementos presiona.
 - **Historial de búsqueda en la app:** incluye información sobre las búsquedas que realizó el usuario en la *app*.
 - **Apps instaladas:** incluye información sobre las apps que están instaladas en el dispositivo del usuario.
 - **Otro contenido elaborado por usuarios:** por ejemplo, biografías de usuarios, notas o respuestas.
 - **Otras acciones:** por ejemplo, información de juegos, opciones de diálogo o 'me gusta'.

8.2. Resumen de beneficios potenciales

Los principales beneficios identificados de esta fuente de datos son:

- **Buen acceso debido a la confianza:** los SDK en *apps* de dispositivos son una manera muy útil de recolectar datos de clientes actuales o en *apps* que ya están bien posicionadas o que tienen buena reputación en las tiendas de apps como PlayStore o App Store, lo cual hace que los usuarios tengan una buena propensión a descargarlas por la confianza que han logrado.
- **Disponibilidad de datos:** el uso extensivo de dispositivos móviles hace que se procesen suficientes datos a ser recolectados. Un SDK puede recoger entre cinco y veinte mil puntos de datos de un solo dispositivo.
- **Cobertura amplia:** como ya se mencionó, Colombia presenta una penetración de apps de servicios financieros del 46,5 %, lo que representa una cobertura favorable. Por otro lado, se necesita tener un SDK por cada tipo de sistema operativo. Los SDK de diferentes sistemas operativos recolectan diferentes puntos de datos que pueden no ser consistentes. Sin embargo, el 89 % de los usuarios colombianos ingresan a internet a través de Android y el 10 % por iOS²⁶; este último ha aumentado 3,5 % en 2022 vs. 2021. Debido a que las EPC corresponden a una población en la base de la pirámide, se espera que la cobertura sea de más del 95 %. Esto supone una ventaja ya que solo con un SDK de Android, se puede garantizar la cobertura y disponibilidad sin necesidad de incurrir en la inversión necesaria para una segunda implementación en iOS.

²⁶ De acuerdo con <https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-colombia-en-el-2021-2022/>

- **Tiempo de respuesta:** de acuerdo con el análisis realizado a diferentes proveedores²⁷ para obtener un resultado, ya sean estos datos crudos, *insights* o *scores*, se identificó que dos de los principales proveedores aseguran que estos se obtienen inmediatamente o en cuestión de segundos.

8.3. Resumen de retos potenciales

Los retos identificados de esta fuente de datos son:

- **Rastreo y compartición de datos:** existen restricciones en que fabricantes como Android o iOS (Sistema operativo móvil de código cerrado) (iPhone Operating System, por sus siglas en inglés) limitan el acceso de datos de parte de los desarrolladores, por ejemplo, Google Play restringe el uso de ciertos permisos de registro de llamadas, a menos que la *app* cumpla con un conjunto particular de requisitos. Por ejemplo: SMS o registro de llamadas. Esto puede impactar en que una aplicación pueda no ser aprobada para ser publicada en la tienda de *apps* del fabricante. Por ejemplo, el PlayStore puede reservarse el derecho a publicar una *app* o inclusive deshabilitar una *app* si no cumple con aspectos de su *Prominent disclosure policy*.
- **Memoria en dispositivos:** al tratarse de código operativo dentro de una aplicación móvil, instalar un SDK requiere espacio de almacenamiento. Ciertos usuarios con restricciones de espacio en sus dispositivos son propensos a eliminar las aplicaciones. Por esto, las organizaciones también son muy cuidadosas en mantener sus aplicaciones ligeras y así evitar desinstalaciones o pérdida de usuarios.
- **Confianza de los usuarios:** si se desea conocer nuevos clientes o aumentar inclusión financiera con prospectos desconocidos, se tiene por obligatoriedad que el usuario descargue una aplicación. Esto en ocasiones representa una barrera de entrada, ya que ciertos usuarios son muy sensibles a ser forzados a descargar una aplicación, dar permisos y recolectar datos con una FI o agregador con el cual no han tenido relación previamente.
- **Esfuerzos de implementación:** el diseño, desarrollo, prueba, despliegue e implementación de una aplicación móvil requiere de una serie de esfuerzos coordinados entre las áreas de producto, TI desarrollo, TI administración de base de datos, analítica, experiencia de usuario, *marketing*, comercial, canales digitales y áreas legales. Lo cual puede hacer que ciertas FI sean desincentivadas de hacer un piloto de cara al esfuerzo e inversión requerida.
- **Especificidad:** siempre y cuando el dispositivo móvil sea el usado por el dueño, y para la actividad económica de la EPC como principal tomador de decisión dentro del micronegocio, los datos pueden ser considerados individuales y específicos. De lo contrario, se pueden tener inconvenientes de especificidad en los datos (similar a los que se tiene con data de Telcos y datos de servicios públicos), ya mencionados.

²⁷ LenddoEFL. <https://lenddoefl.com/scoring>. Begin: <https://www.begini.co/device-data/>

- **Requerimientos de red para transmisión de datos desde el dispositivo:** si bien se identificó que los proveedores afirman que los resultados se pueden obtener inmediatamente después de que se recolectan los datos del dispositivo, existen aspectos como velocidad de red del dispositivo, zona geográfica, tipo de conexión, latencia, así como cantidad de datos en el dispositivo móvil que pueden incidir en el tiempo de respuesta por parte de los proveedores. Esto le puede restar dinamismo a una implementación en zonas rurales o de pobre conectividad de red.

9. Datos de e-mail

9.1. Generalidades

El uso de datos de correo se ha extendido debido a que ingresar un correo electrónico o e-mail es parte de los datos mínimos requeridos a los solicitantes de crédito. Este mecanismo evalúa la consistencia de los datos del usuario.

Diferentes organizaciones han estudiado los patrones de uso del correo electrónico y se ha mostrado que pueden indicar el grado de actividad y compromiso de una empresa, así como también la calidad de sus relaciones comerciales. También es posible que el análisis de la frecuencia y el contenido de los correos electrónicos pueden proporcionar información sobre la salud financiera de un negocio, así como también sobre sus posibles necesidades de financiamiento.

La empresa Lenddo utilizó datos y metadatos de correos electrónicos, junto con otras fuentes de información, para crear scores de crédito donde se analizaron los patrones de uso de correo electrónico de los usuarios, incluyendo la frecuencia de uso, la calidad de los contactos y el lenguaje utilizado en los correos electrónicos. Estos datos se combinaron con información de otras fuentes, como las redes sociales, para crear un perfil de crédito del usuario.

De manera similar, la compañía de tecnología y analítica ZestFinance provee servicios de analítica avanzada para IF, la cual utiliza aprendizaje automático (o Machine Learning) para hacer evaluaciones de riesgo crediticio, donde se menciona que podría determinar si una persona paga las facturas de servicios públicos a tiempo con regularidad, en función de si envía correos electrónicos de manera constante desde una cuenta de Gmail o Yahoo, en un momento determinado del día o de la semana bajo el supuesto de que está muy involucrado en sus finanzas personales (Popper, 2014).

Existen también otras experiencias internacionales directamente de empresas de préstamos en línea como Kreditech, que establece cuando utiliza diversos tipos de datos, incluyendo información de correo electrónico y otras fuentes digitales, para evaluar la calidad crediticia de los solicitantes de préstamos, donde sus algoritmos analizan los patrones de uso de correo electrónico de los usuarios para evaluar su fiabilidad en cuanto a repago de los préstamos (Goodman, 2014).

Al igual que ocurre con los datos de teléfono, el correo electrónico es registrado por el posible cliente en su solicitud y este tiene una doble finalidad:

- Verificar acceso a la aplicación.
- Determinar factores de riesgo de fraude mediante el uso y la antigüedad.

La tabla 10-1 presenta una serie de variables (también llamada “variable estadística”), (así como su función), obtenidas del correo electrónico y usadas en un banco latinoamericano. De tal modo, se observa que la función principal del modelo es identificar aplicaciones fraudulentas o aplicantes, que tendrían una tendencia a cometer abuso con el uso del crédito. Sin embargo, no se han identificado documentos o estudios que relacionen la probabilidad de incumplimiento con las características del uso del correo electrónico en propietarios de micronegocios.

*Tabla 10-1. Variables seleccionadas a evaluar en datos de e-mail**

Función	Subcampo	Variable	Acción
Acceso a internet	Navegador	Safari, chrome, mozilla, etc	Avanza
	Navegador	TOR, 12P, freenet, etc	Declina
Seguridad	Dominio	Gmail, yahoo, icloud, etc	Avanza
	Dominio	Inexistente	Declina
	Dominio	Riesgo bajo	Avanza
	Dominio	Riesgo medio	Revisa
	Dominio	Riesgo alto	Declina
Nivel de confianza	Antigüedad	>365 días	Avanza
	Antigüedad	<365 días	Revisa
	Antigüedad	<60 días	Declina
	Fraude confirmado	Sin marca	Avanza
	Fraude confirmado	Con marca	Declina
	Válido	Si	Avanza
	Válido	No	Declina
	Sintaxis	Sin patrón de riesgo	Avanza
	Sintaxis	Con patrón de riesgo	Declina
	Redes sociales	Ligado a cuentas sin riesgo	Avanza
	Redes sociales	Ligado a cuentas de riesgo medio	Revisa
	Redes sociales	Ligado a cuentas de alto riesgo	Declina
	Dominio	Riesgo bajo	Avanza
	Dominio	Riesgo medio	Revisa
	Dominio	Riesgo alto	Declina
	Actividad	Bandeja de entrada saturada	Declina
	Actividad	Cuenta inactiva / cerrada	Declina
	Actividad	Frecuencia de uso regular	Avanza
	Actividad	Frecuencia de uso nula	Declina
	Cuenta en riesgo	Asociada a otros factores de riesgo	Declina
	Cuenta en riesgo	Posible cuenta comprometida	Declina

Tabla 10-1. Variables seleccionadas a evaluar en datos de e-mail* (continuación)

Función	Subcampo	Variable	Acción
Nivel de confianza	Cuenta en riesgo	Marcada como fraude por alguna tienda o servicio	Declina
	Cuenta en riesgo	Marcada como fraude por algún medio de pago	Declina
	Cuenta en riesgo	Marcada con múltiples reclamaciones	Declina
	Cuenta en riesgo	Marcada con fraude no atribuible a solicitante	Revisa
	Dirección IP	Regular	Avanza
	Dirección IP	Irregular	Revisa
	Dirección IP	Antecedentes negativos	Declina
	Dirección IP	No encontrada	Declina

*Tabla ilustrativa y no exhaustiva.

Fuente: Creación propia basada en experiencia de trabajo en la implementación de un *scoring* alternativo para un banco mexicano.

En un estudio que buscaba mejorar la calificación crediticia con datos alternativos (Viani B. Djeundje, 2020) se diseñó un modelo basado en volumen y contenido de correos enviados y recibidos²⁸. Se intentó demostrar que la probabilidad de incumplimiento se puede asociar con la fracción de correos electrónicos enviados en ciertos horarios. Sin embargo, como señala el estudio de EFL, el tiempo requerido para recopilar estos datos es exigente, por lo que no se pueden tomar decisiones rápidas cuando se analizan cuerpos completos de correos electrónicos, reforzando el uso de estos datos como mecanismo para detección de fraude, y no como un elemento acertado para predecir el comportamiento crediticio (Center for Financial Inclusion at Accion e Institute of International Finance, 2018).

9.2. Resumen de beneficios en uso de datos de e-mail

- **Permite detección de fraude con asertividad:** distintos puntos de uso del correo evalúan y perfilan a solicitantes con un riesgo de fraude, permitiendo a la institución financiera identificar tales solicitudes e investigar más.
- **Cobertura amplia:** el creciente uso del correo ha derivado en que un volumen alto de la población disponga de al menos uno. También el fácil acceso, y como requisito creciente en distintos escenarios ha contribuido a incentivar su uso. Dicho uso deriva en la producción de datos que se pueden rastrear.

9.3. Resumen de retos en uso de datos de e-mail

- Para el ecosistema en el que se evalúa implementar existen dificultades en la estandarización de variables y resultados de las evaluaciones, ya que el uso del correo electrónico no está ligado necesariamente a las actividades empresariales o de negocio.

²⁸ El estudio se centró en diseñar un modelo que combina los datos sociodemográficos, los resultados del uso del correo electrónico y los psicométricos para la calificación del crédito. Los datos de uso del correo electrónico contenían el número de correos enviados y cuáles fracciones de estos se enviaron en diferentes días. Los datos sociodemográficos incluían la edad, el sexo y el número de personas a cargo, entre otros factores. No se concluyó que el uso de los datos del correo mejorara la precisión, aunque sí se encontraron evidencias, los cuales apuntan a que la evaluación de los datos de correo puede contribuir a evaluar la probabilidad de impago.

- La puesta en marcha de un modelo de *scoring* depende de la evaluación de un volumen alto de datos, lo que incrementa las necesidades tecnológicas y de almacenamiento que puedan proveer un resultado rápido.
- No se identificaron experiencias concretas de su uso para construcción de modelos de riesgo de crédito en micronegocios, tan solo algunos, pero en créditos personales de bajo monto y plazo.
- Uso extendido y probado para detección de fraude, pero los casos de éxito en la producción de *scoring* de crédito son limitados, y siempre van acompañados de otras fuentes de datos como redes sociales, datos de Telcos o metadatos de exploración web, por lo que se supondría que esta fuente de datos de manera independiente podría tener poca precisión.
- De manera similar a los datos de redes sociales pueden resultar relevantes para identificar voluntad de pago, sin embargo, no son necesariamente precisos para medir la solvencia o capacidad de pago, ya que solo pueden hacer una ‘inferencia’ acerca del comportamiento financiero, los ingresos y hábitos de consumo, pero poca información relevante sobre las deudas o el historial de pagos de los usuarios.
- Una consideración importante para usar datos de *e-mail* radica en que usuarios pueden inducir la producción de tráfico en sus correos electrónicos, por ejemplo suscripciones a *newsletter* de portales de finanzas o economía, múltiples alertas de precios de aerolíneas, cambiar notificaciones de consumos para que tanto al dispositivo como al correo (p. ej. Uber) para producir una “huella digital artificial”, es decir, una huella digital que no es resultado natural del uso habitual de la misma red social y esto puede influir en los resultados que arroje un modelo de *score* de crédito.

10. Datos de transacciones públicas

10.1. Generalidades

En esta investigación se definen los datos derivados de las transacciones públicas como aquellos resultados de actividades de las EPC con lo público, como por ejemplo pagos a compañías de servicios públicos, aportes a salud, información fiscal, propiedades y subsidios gubernamentales; es decir, todos aquellos datos que producen información transaccional de las obligaciones de las EPC para poder operar, pero que son producidos por entidades distintas a las financieras. Esta situación se hace porque la penetración de estos flujos es mayor a la bancaria, y porque al ser provenientes actividades cotidianas de las EPC con lo público, proveen un entendimiento distinto del usuario en comparación con los datos bancarios que son primordialmente actividades cotidianas con lo privado.²⁹

En el caso de Colombia, estos datos se consideran como Open Finance, que similar al *open banking*, explota los datos transaccionales provenientes de entidades no estrictamente financieras, pero que procesan información financiera de sus usuarios.

²⁹ Esta separación se realiza a petición de los patrocinadores del proyecto y en acuerdo con el equipo consultor

Los “datos transaccionales” se refieren a los registros de las actividades comerciales entre una empresa proveedora (ya sea privada o pública) y sus clientes. Por lo general, incluyen información relacionada con los pagos (datos de flujo de efectivo) e información no relacionada con dinero (datos que no son de flujo de efectivo). Las tendencias de comportamiento emanadas por el análisis de esta información se pueden utilizar para evaluar los ingresos y el estado financiero más reciente de una mipyme. Al mismo tiempo, la información no financiera derivada puede producir *insights* para predecir la solvencia de una empresa, así como sus hábitos y comportamientos respecto a los pagos.

Los datos de transacciones públicas en la toma de decisiones y la producción de términos de crédito pueden aumentar tanto el tamaño de la población que se puede evaluar, como la economía y sanidad de los préstamos otorgados, sin cambios sustanciales en la evaluación del riesgo.

Similar a los datos de *open banking*, estos pueden ser recabados y entregados a las IF por medio de agregadores *fintech*, que forman alianzas con los distintos proveedores. Entre los proveedores de datos transaccionales no tradicionales principales se encuentran servicios públicos; aportes a pensión, salud y riesgos; información fiscal; subsidios; y propiedades y catastro.

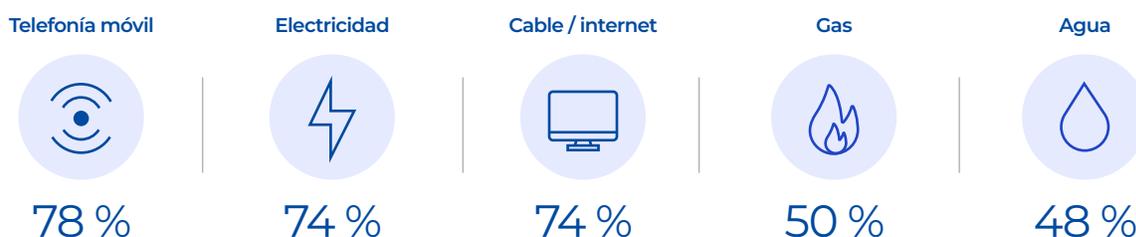
10.1.1. Servicios públicos

La posibilidad del uso de esta fuente depende fundamentalmente del acceso a ella. Las principales formas de acceso son:

- a. Acuerdo institucional entre la empresa de servicios públicos con otra organización que puede ser un agregador, una agencia/firma de analítica de datos, una central de información o una entidad financiera.
- b. Acceso directo *opt-in* de parte del usuario, que haga uso de la plataforma de la entidad prestadora del servicio donde da permisos para que un agregador, agencia/firma de analítica de datos o entidad financiera acceda a los datos de su cuenta.

En una encuesta en 2019, se descubrió que aproximadamente el 92 % de los solicitantes de crédito estadounidenses tienen al menos un servicio público o relación de pago de telecomunicaciones en su nombre, y el 59 % está dispuesto a compartir estos datos con una entidad de crédito para mejorar su acceso al crédito y otros servicios. Los consumidores a quienes previamente se les había negado un préstamo estaban aún más dispuestos (66 %) a hacerlo si esto aumenta sus posibilidades de acceder al crédito (Urjanet, 2019) (figura 10-1).

Figura 10-1. Penetración del consumidor de las relaciones de pago de servicios públicos y telecomunicaciones



Distintas empresas Fintech en el mundo han comenzado a explorar el perfilamiento de clientes a partir de su historial de pagos de servicios públicos, siendo una de ellas la empresa Ágata, en Colombia.

Ágata tiene una alianza estratégica establecida por la Empresa de Telecomunicaciones de Bogotá (ETB); el Grupo de Energía de Bogotá, la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB), la Unidad Administrativa Especial de Catastro Distrital y la Secretaría Distrital de Planeación³⁰; esta alianza, provee riqueza a los datos y propicia una solución con cobertura y disponibilidad de datos amplias. Este proveedor se posiciona como un aliado de las instituciones financieras, en colaboración con la construcción de modelos de riesgo de crédito usando estos datos en conjunto con otras fuentes

10.1.2. Aportes a salud, pensión, riesgos

Similar al caso de los servicios públicos, el acceso a los datos de aportes a salud, pensión y riesgos depende de los acuerdos bilaterales que se forman entre los jugadores Fintech y las instituciones que disponen de la información.

En Colombia, las leyes 1266 de 2008 y 1581 de 2012, que cubren aspectos clave de la compartición de datos personales entre instituciones, ha facilitado que este flujo de información se institucionalice y sea posible producir scores alternativos con miras a incrementar el acceso al crédito. Sin embargo, los decretos recientes relacionados con Open Finance aún no hacen obligatorio para este tipo de instituciones implementar API de manera obligatoria.

Las soluciones existentes se encuentran en un estado naciente, pero se han montado pilotos importantes hacia el crecimiento de estas soluciones. Estas soluciones se apalancan en la información en tiempo real de los pagos de seguridad social e información no tradicional para conocer los ingresos y el empleador que reporta una persona natural (empleado) o un trabajador independiente (mipyme). En el caso concreto de Colombia, se ha identificado a.

El reto en este tipo de soluciones se encuentra en la cobertura, ya que las mipymes tradicionalmente se encuentran en un sector informal que no reporta aportaciones. Por el contrario, la cobertura se extiende a personas físicas y jurídicas con aportaciones a salud y pensiones. Sin embargo, tienen la ventaja de tomar en cuenta los ingresos variables y estimar una proyección de ingresos en doce meses.

Estas soluciones digitales usan en Colombia cuatro fuentes de datos:



³⁰ <https://agatadata.com/nosotros.html>

Estas fuentes de datos se agregan por medio de algoritmos desarrollados por los fintech y se entregan al consumidor de datos para su uso (figura 10-2).

Figura 10-2. Flujo de datos de aportaciones



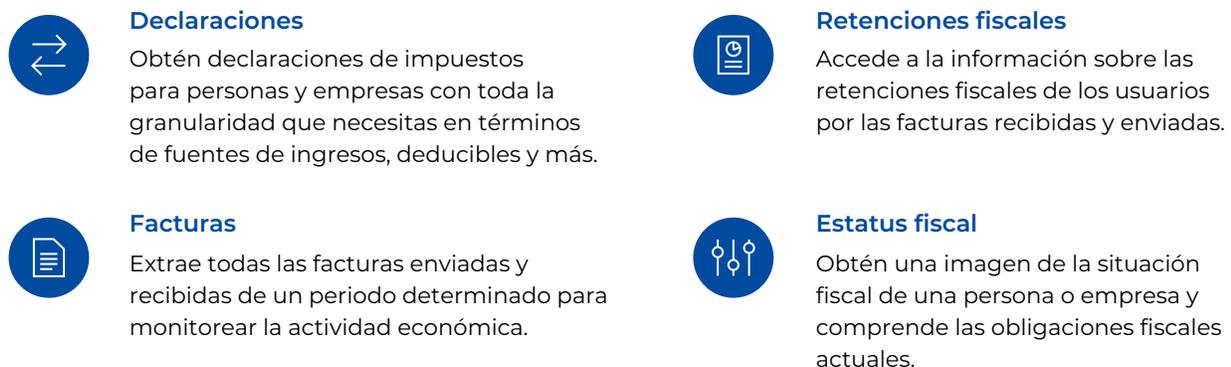
Fuente: Mareigua.

En el caso concreto de Colombia, la Fintech Mareigua destaca por haber construido relaciones con los proveedores principales de esta fuente de datos y la infraestructura necesaria para facilitarlos a distintos consumidores, incluyendo instituciones financieras.

10.1.3. Información fiscal

Como ya se mencionó, similar a como se realiza con los datos de *open banking*, estos ya pueden ser recabados y entregados a las IF por medio de agregadores fintech; y se realizan al acceder a los datos y productos fiscales de la autoridad fiscal de un país, que para Colombia sería la Dirección de Impuestos y Aduanas Nacionales (DIAN) (figura 10-3).

Figura 10-3. Tipos de datos fiscales obtenidos a través de agregadores (Belvo) de open banking



Fuente:

El proveedor asegura que con los datos extraídos puede obtener detalles sobre la situación fiscal de una persona o empresa incluyendo patrimonio, ingresos, impuestos a pagar y deducciones.

Se podrían realizar estimaciones de ingresos al obtener información de facturación electrónica tales como: remitente, receptor, importe y moneda, datos de la factura y el tipo de pago al que corresponde.

Respecto a los costos de este servicio, tal y como se describió en la tabla 10-3 de *open banking* para el caso de los datos fiscales se cobra una cuota fija por cada consulta realizada, que puede costar alrededor de 0,1 USD (480 COP).

Otra manera de obtener los datos fiscales en teoría podría ser si la DIAN comparta la información fiscal de los micronegocios para que sean usados por las entidades financieras o agencias/firmas de analítica de datos, y de esta manera realizar análisis de crédito a través de acuerdos interadministrativos, siempre y cuando se tenga la autorización previa por parte del representante legal de cada uno de los micronegocios a ser consultados.

10.1.4. Subsidios

Los subsidios entregados por los gobiernos a familias y personas que cumplen con ciertas condiciones es una fuente de datos que podría incluirse en la producción de un *score* crediticio, ya que forma parte de los ingresos de la gente. En el caso de Colombia, actualmente estos subsidios se han entregado por medios digitales y formaron parte del avance que se ha hecho en materia de inclusión financiera.³¹

Concretamente, el ingreso solidario benefició a tres millones de personas, entre los que se encuentran microempresarios y dueños de micronegocios con necesidades financieras. Los subsidios fueron entregados por medios digitales: (a) en cuentas bancarias para aquellos que ya tenían una cuenta, y (b) se incentivó la apertura de cuentas e Nequi y Daviplata para aquellos que no tenían una cuenta bancaria.

Como se mencionó anteriormente, estimaciones recientes indican que Nequi y Daviplata tienen más de 20 millones y 22 millones de usuarios, respectivamente³² y parte del crecimiento acelerado se ha debido a la entrega de subsidios por este medio (~est. 1,2 millones de usuarios únicos nuevos³³).

A pesar de que, a partir de 2023, el ingreso solidario dejará de entregarse, otros programas sociales con traspaso de subsidios continúan y se incluirán otros. Se espera que los depósitos continúen haciéndose por medios digitales y que puedan ser usados por medio del *open banking*.

³¹ Al final de 2022 el índice de inclusión financiera (acceso) era mayor al 90 % y se estima que más de 34 millones de adultos tienen un producto financiero de depósitos.

³² Con datos recabados de Data.ai el 1 de marzo de 2023 y contando con 1,2 de descargas por usuario (Data.ai es un rastreador de aplicaciones que, entre otros datos, mide el número de descargas y usuarios activos de aplicaciones móviles).

³³ Estimación propia basada en datos de cuentas nuevas 2021 y 2022, y reportes anuales de Bancolombia y Davivienda (empresas de las cuales Nequi y Daviplata son subsidiarias).

Otra manera de obtener los datos fiscales, en teoría, podría ser si la DIAN comparta la información fiscal de los micronegocios para que sean usados por parte de las entidades financieras o agencias/firmas de analítica de datos para realizar análisis de crédito a través de acuerdos interadministrativos, siempre y cuando tenga la autorización previa del representante legal de cada uno de los micronegocios a ser consultados.

10.1.5. Información de procesos judiciales

La firma LexisNexis³⁴ asegura que ha usado datos de registros públicos para encontrar *insights* predictivos sobre pequeñas empresas y sus propietarios. Identificó que el desempeño de pago correlacionaba con la presencia o ausencia de gravámenes o sentencias, buena reputación con la secretaria de Estado, títulos de propiedad, y dueños de negocios que tienen condenas por delitos graves (LexisNexis, 2018).

10.2. Experiencia que combina diferentes fuentes de datos públicos en EPC en Bogotá, Colombia. El caso Ágata y la SDDE

Ágata³⁵, es una agencia analítica de datos que provee soluciones tanto de analítica como de SaaS, que nace con la alianza estratégica establecida por la Empresa de Telecomunicaciones de Bogotá (ETB); el Grupo de Energía de Bogotá, la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB), la Unidad Administrativa Especial de Catastro Distrital y la Secretaría Distrital de Planeación.

Actualmente está trabajando en diferentes etapas de un gran programa con diferentes aliados financieros tanto de la banca tradicional como Fintech e IMF, para ofrecer microcrédito y créditos de bajo monto de manera digital a dueños de pequeños negocios y producto de estos esfuerzos, en conjunto con la Alcaldía Mayor de Bogotá a través, de la Secretaría Distrital de Desarrollo Económico (SDDE) y Ágata crearon la plataforma “Crédito Peso a Peso”, que utiliza los datos públicos que se han recopilado de estos dueños de pequeños negocios desde las distintas instituciones que componen la Alcaldía y, con la debida autorización sobre el uso de esta información, se transforman en calificaciones alternativas del riesgo (Secretaría Distrital de Desarrollo Económico, 2023).

De esta manera, se aprovechan los acuerdos interinstitucionales en esta alianza para beneficiar a los datos de la ciudad que son un activo subutilizado, para poderlos evaluar con metodologías de puntaje alternativo.

Entre los datos públicos que usan, se destacan los siguientes:

- Fuentes de registro social.
- Fuentes de ciudades, tales como servicios públicos y datos de circulación.
- Base maestra de la ciudad, con datos de población en condición de vulnerabilidad, e información del Sisbén, entre otros.

³⁴ LexisNexis® Risk Solutions brinda soluciones y herramientas de decisión que combinan contenido público y específico de la industria para evaluar y predecir el riesgo.

³⁵ <https://agatadata.com/nosotros.html>

- Datos relacionados con tareas de regulación, por ejemplo, Cámara de Comercio y bomberos, entre otros.
- Información de ruralidad.
- Datos de producidos directamente en campo resultado de levantamiento interno.

En la etapa anterior del piloto, se colocaron 177 créditos y en 2023 se espera llegar a los diez mil haciendo unas mejoras a la propuesta de valor de producto y añadiendo garantías. En el 2022 finalizaron la producción de los dos primeros modelos de score de crédito alternativo usando las múltiples fuentes de datos públicos antes mencionadas y durante el año 2023 se aspira a operar con cuatro modelos³⁶.

Como resultados parciales se han obtenido tasas de aprobación de entre el 10 % y el 12 % para la banca y las IMF, mientras que alrededor del 50 % para los fintech. Se espera con el monitoreo de los modelos y las implementaciones de las propuestas de valor en el producto y garantías que la tasa de aprobación se estabilice entre el 24 % y 33 %.

10.3. Resumen de beneficios potenciales de datos de transacciones públicas

- Se podría tener una interesante cobertura y disponibilidad de datos de servicios públicos, en especial en zonas urbanas con los correspondientes acuerdos interinstitucionales con agregadores o entidades financieras.
- Parece existir una voluntad por parte de los usuarios de compartir información de su cuenta, si esto implica beneficios de acceso a crédito.
- Puede contener un gran volumen de datos tanto financieros como no financieros, que además de la implementación de modelos de crédito, adicionalmente conduce a la construcción de productos y propuestas de valor considerando aspectos como actividad económica específica, referenciación geográfica y nivel de condición de vulnerabilidad, entre otros.
- El nivel de especificidad de ciertas fuentes públicas de ciudades y ruralidad puede hacer más asertiva la focalización de la propuesta de valor para dar acceso al crédito a quien más lo necesita, de manera que los modelos de crédito consideren esa propensión en sus calificaciones.
- El uso de datos de servicios públicos tiene más trayectoria y experiencia internacional para casos de uso de riesgo de crédito, debido principalmente a la cobertura en especial en zonas urbanas.
- Las inversiones requeridas para hacer una implementación y consulta recurrente de datos fiscales en general son bajas y pueden ser atractivas a pesar de su posible baja cobertura.

10.4. Resumen de retos potenciales en datos de transacciones públicas

A pesar de tener una cobertura general más amplia, así como disponibilidad, debido a las alianzas que se han tendido, en el caso de las mipymes la cobertura disminuye debido a los siguientes factores:

³⁶ Obtenido de reuniones y entrevistas con equipo de Ágata.

- Existe un reto sobre la identificación, si se trata de un predio para uso comercial de la EPC o el domicilio de la persona que lidera ese micronegocio o emprendimiento, que permita discernir o calificar esa actividad económica.
- En el caso de aportaciones o pagos de impuestos, un porcentaje importante de la población objetivo es informal, por lo que no se incluyen en los servicios de los agregadores.
- Es frecuente encontrar inconsistencias entre los titulares de una cuenta o contrato o referencia de pago. Generalmente no hay correspondencia entre el nombre y cédula de un titular, en especial si se trata de un inmueble en arriendo o con alguna inconsistencia catastral.
- Por lo general, se tiene una cuenta o contrato o referencia de pago por predio, esto dificulta el principio de especificidad al pretender evaluar a un solo individuo, en este caso el titular de la actividad económica EPC cuando en el domicilio habitan varios.
- Existen varios retos de cobertura; por un lado, que no de todos los usuarios se tendrán datos de todas las subfuentes, ya sea porque se encuentren con diferentes empresas de servicio público o por las inconsistencias de nombres de titulares por lo que se requerirán tratamientos de datos para considerar esos datos perdidos o faltantes.
- Se espera que un grupo importante del segmento EPC no haga pagos a través de plataformas digitales.
- La modalidad de *opt-in* para servicios públicos no ha sido masificada en Colombia, como en otros países, por lo que esta alternativa no sería viable en el mediano plazo a pesar de la voluntad expresada en estudios de otros países.
- La información de aportes se puede tomar como datos crudos de operadores de información PILA, sin embargo, debido a la informalidad del segmento EPC no se espera que tenga mucha cobertura.
- De la misma manera, la información fiscal se puede tomar como datos crudos de agregadores o a través de acuerdos interinstitucionales, en cuanto a facturación electrónica o información de declaraciones, sin embargo, debido a la informalidad del segmento EPC no se espera que tenga mucha cobertura.

Los datos de transacciones públicas pueden ser muy ricos, en especial si se manejan diferentes subconjuntos, sin embargo, la obtención de la autorización de uso de los datos por parte del dueño del micronegocio para cada producción de los datos (p. ej. alcaldía, empresa de servicios públicos, caja de compensación, etc.) resulta complicada en el proceso.

Los datos de o pagos a parafiscales tienen poca aplicabilidad en el segmento EPC, debido a la alta formalidad que maneja este. De acuerdo con la Fintech Mareigua, los micronegocios o personas naturales con actividad empresarial tienen poco interés en hacer aportes al régimen contributivo de salud, en primera instancia porque les produce un costo y adicionalmente existe la sensación de que pierden beneficios, por ejemplo, subsidios del Sisbén. Por lo anterior, al considerar el uso de fuentes de transacciones públicas que el segmento de micronegocios tiene poco interés en formalizarse (probablemente por la falta de estímulos para ello), por lo que es importante tener en cuenta que posiblemente nunca lleguen a tener personería jurídica.

10.5. Principales proveedores

Para la fuente de datos transaccionales también se realiza una preidentificación de soluciones a partir de su oferta de valor (tabla 10-1). En el análisis se incluyeron aquellas firmas y agencias que tienen como oferta de valor la agregación de datos para consumo por entidades financieras, o que tienen una solución directa para *scoring* alternativo.

Tabla 10-2. Preidentificación de proveedores de datos transaccionales

Compañía	HQ	Categoría	Sitio web	Año de fundación	Experiencia Colombia	Comentarios
Ágata Data ★	Colombia	Agregación y analítica de datos	www.agatadata.com	2021	Sí	Agregador de datos con convenios de múltiples transacciones públicas en Bogotá principalmente.
Mareigua	Colombia	Transacciones públicas	www.mareigua.co	2018	Sí	Agregador de datos con convenios aportes, fiscal
Belvo ★	México	Transacciones públicas	belvo.com	2019	Sí	Plataforma API de finanzas abiertas.

★ Proveedores preseleccionados.

Fuente: elaboración propia, basada en investigación previa, bases de datos y asociaciones Fintech.

10.6. Análisis de propuestas comerciales

Las ofertas comerciales en datos de transacciones públicas dependen del proyecto a implementar, al ser provistos por agencias de analítica de datos, o por firmas en proceso de consolidación y que están evolucionando su propio modelo de ingresos. En esta sección se detalla la propuesta de un proveedor de Open Finance que tiene dentro de su abanico de datos transacciones públicas en Colombia (por ejemplo, subsidios entregados digitalmente y datos fiscales).

10.6.1. Inversión estimada

De manera similar a datos de psicometría (sección 5.9), la inversión en soluciones de transacciones está asociada a las etapas del proyecto de implementación y requerimientos mencionados anteriormente. También se identificó que existen tres momentos de inversión: instalación inicial, costo recurrente y servicios adicionales. A continuación, se describe cada uno de ellos, con unos rangos de costo.

a. Instalación inicial

Los proveedores consultados establecen una cuota por integración de las API que dependen del nivel de complejidad de instalación. Como se describe en la tabla 10-2, el precio varía entre 1.000 USD (~4,8 millones COP) a 2.000 USD (~10 millones COP).

Tabla 10-3. Escenarios de costos de instalación de datos de transacciones públicas

Concepto	Complejidad baja	Complejidad alta
Costo por integración a API de agregación (completa)	1.000 USD	2.000 USD

* Estimaciones basadas en ofertas comerciales vigentes de proveedores a febrero de 2023.

b. Costos recurrentes

Los proveedores de *open banking*, tienen cuota mensual que provee acceso a la plataforma, al servicio al cliente, y otros costos de operación que varía entre 500 USD (~2,5 millones COP) a 1.000 USD (~4,8 millones COP). El precio se establece con relación al tamaño de la organización contratante y del volumen de uso esperado (tabla 10-3).

Tabla 10-4. Escenarios de costos recurrentes mensuales de datos de transacciones públicas*

Concepto	Mínimo	Máximo	Recurrencia
Suscripción (servicio al cliente, gestión de cuenta, ingeniería)	500 USD	1.000 USD	Mensual

Fuente: Estimaciones basadas en ofertas comerciales vigentes de proveedores a febrero de 2023.

c. Costos por consulta

Finalmente, se cobra una cuota fija por cada consulta realizada, que puede ir desde 0,1 USD (480 COP) a 0,7 USD (3.400 COP), que incluye la revisión completa de los puntos de datos bancarios, fiscales, enriquecimiento de la consulta y categorización (evaluación).

En total, el costo estimado anual, considerando mil consultas mensuales es de 15.400 USD (~7.4 millones COP), o diez mil consultas mensuales que es de 95.000 USD (~457 millones COP) (tabla 10-4).

Tabla 10-5. Escenarios de costos por consulta de datos de transacciones públicas

Concepto	Mínimo	Máximo	Recurrencia
Agregación fiscal	-	0,10 USD	Por revisión
Agregación bancaria (para identificar subsidios)	-	0,30 USD	Por revisión

Fuente: estimaciones basadas en ofertas comerciales vigentes de proveedores a febrero de 2023.

11. Entorno normativo

En esta sección se discuten las principales normas aplicables a las fuentes de datos seleccionadas. Una de las razones por las cuales las IF vigiladas por la Superintendencia Financiera de Colombia no han implementado este tipo de modelos de manera generalizada se relaciona con la regulación, la cual pretende evitar que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla con sus obligaciones³⁷ (Superintendencia Financiera de Colombia). Esto obliga a manejar altos estándares que dificulten a las entidades utilizar estos nuevos modelos, los cuales conllevan a un nivel de riesgo alto (Ustariz y Abogados, 2020).

Por otro lado, la Superintendencia de la Economía Solidaria, a través de la Circular Básica Contable, Título IV, Capítulo II, indica que las organizaciones solidarias vigiladas deben implementar el Sistema de Administración de Riesgo de Crédito “con el propósito de identificar, medir, controlar y monitorear el riesgo de crédito al cual se encuentran expuestas en el desarrollo de su proceso de otorgamiento”. Asimismo, señala que las entidades deben adoptar metodologías o criterios que les permitan evaluar el perfil deudor del asociado y su capacidad de pago³⁸, durante la etapa de otorgamiento de crédito.

Estas entidades tienen como referencia principal la proporcionada por centrales de riesgo (Datacrédito y Transunión), además es conocido que se usan procedimientos y metodologías tradicionales que producen demoras en tiempos y que dejan la decisión de otorgamiento de manera subjetiva entre clientes con condiciones similares al carecer de modelos estadísticos. La Supersolidaria ha reconocido que la cartera de crédito que se supervisa ha sufrido deterioro, por lo que ha emitido nuevas circulares y anexos, los cuales implican que las entidades tengan sistemas de medición más estrictos. Es por esto, que la adopción de nuevos modelos de *scoring* alternativo beneficiaría a las entidades solidarias al proveerles herramientas y metodologías para colocar créditos sanos con mayor eficiencia.

Sin embargo, en 2020 se dio un primer paso hacia flexibilizar la regulación gracias al Programa de Acompañamiento a Deudores³⁹ (PAD), con el cual se permitió que los deudores redefinieran las condiciones de sus créditos sin que aumentara la tasa de interés pactada inicialmente. La flexibilización consistió en que tanto para el proceso de otorgamiento de nuevos créditos como para la redefinición de los mismos, las entidades vigiladas podían hasta el 30 de junio de 2021, establecer procedimientos que emplearan información alternativa que les permitiera obtener una percepción razonable y objetiva sobre la capacidad real o potencial de pago del deudor y que reconociera variables adicionales sobre la reactivación futura del sector económico donde se desempeña y su capacidad de generación de ingresos (Ustariz y Abogados, 2020).

³⁷ En Colombia está establecida una solvencia básica del 9 %, pero en la práctica se ha mantenido por encima del 16 %, debido a una postura cuidadosa respecto al capital por riesgo de mercado, coeficiente de cobertura de liquidez y de financiación estable neta.

³⁸ De acuerdo con la normativa emitida por la Supersolidaria en 2020 y actualizada en 2022.

³⁹ Según la Asobancaria y la Superfinanciera, 2,2 millones de deudores redefinieron las condiciones de los créditos por un valor aproximado de 36.000 millones COP. Cerca del 85 % de la cartera está normal, 5 % están en mora y 10 % siguieron en programas de apoyo

De manera similar, el 25 de julio de 2022, se expidió el Decreto 1297, el cual tiene el objetivo de *promover la competencia, la inclusión y la eficiencia en la prestación de productos y servicios*. Esto, mediante permitir a las entidades financieras vigiladas perfilar mejor a los usuarios, y desarrollar estrategias y alianzas con entidades de otros sectores. Gracias a esta nueva vía de regulación:

- Se definen las reglas aplicables al intercambio de datos en entidades financieras, protegiendo al consumidor.
- Se tiene en cuenta la administración de plataformas digitales y la prestación de servicios digitales.
- Se reglamenta la iniciación de pagos.

Las implicaciones del Decreto 1297 del 2022 se resumen en la siguiente sección:

11.1. Entorno normativo de *open banking* en Colombia

El Decreto 1297 supone un avance en la regulación del modelo de *open banking* en Colombia. Define con claridad los jugadores que podrán operar bajo estas nuevas normativas y establece las reglas de funcionamiento que deberán cumplirse cuando se instaure de manera definitiva.

Un aspecto para destacar sobre este decreto es que su entrada en vigencia será al cumplirse doce meses de su expedición. En este tiempo, las entidades financieras participantes deberán estandarizar sus protocolos para la compartición de datos.

Existen tres ejes clave: *open banking*, *banking as a service*, e iniciación de pagos. Este análisis se enfoca únicamente en *open banking*, pero se mencionan aspectos relevantes de los otros dos con el fin de proveer claridad en el alcance.

11.1.1. Iniciación de pagos

Los servicios de iniciación de pagos podrán ser realizados por los establecimientos de crédito, las sociedades especializadas en depósitos y pagos electrónicos (SEDPEs), las entidades administradoras de sistemas de pago de bajo valor y por sociedades no vigiladas por la Superintendencia Financiera de Colombia (SFC).

Algunas normas de funcionamiento incluyen la obligatoriedad de la autorización previa por parte del ordenante, autenticarlo en todos los casos y no solicitar claves o contraseñas. No obstante, se espera mayor claridad sobre algunos de estos términos a través de circulares por parte de la SFC que permitan guiar de mejor manera la implementación de la iniciación de pagos instantáneos y de bajo valor.

Los iniciadores de pagos pueden ser cualquiera de los actores supervisados por la SFC (establecimientos de crédito, aseguradores, administradores de pensiones y cesantías, fiduciarios, intermediarios de valores, portafolios de inversión y conglomerados financieros), pero también por entidades que no pertenecen al sistema financiero, por lo que se incrementa la oferta de pagos digitales (con un enfoque importante en comercio electrónico).

11.1.2. *Open Finance (open banking)*

Implicaciones

- a. Permite la transferencia de los datos del consumidor entre entidades financieras, lo que incrementa la oferta de servicios.
- b. Garantiza un manejo adecuado sobre la información del consumidor, y al mismo tiempo busca diversificar la oferta de servicios.

En esencia, el decreto busca que la información anterior, la cual se quedaba entre usuario y un primer proveedor de servicios financieros, sea compartida con otros jugadores del ecosistema, se abre la barrera, y que esa información la usen muchos oferentes financieros y puedan diseñar nuevos productos (incluyendo crediticios) más diversos, más baratos y con mayor acceso.

Hay aspectos que abren la puerta a que los modelos basados en infraestructuras de *open banking* sean viables en Colombia:

- El documento destaca que las entidades supervisadas por la Superintendencia Financiera de Colombia (SFC) pueden comercializar el uso, almacenamiento y circulación de los datos personales, siempre que tengan la respectiva autorización expresa del titular de los datos.
- También afirma que las entidades supervisadas podrán ofrecer en sus canales virtuales y presenciales productos y servicios de terceros, siempre que este disponga de la autorización en sus operaciones de conexión.

En este sentido, el decreto no plantea ninguna modificación de las actuales normas relacionadas con la protección de datos. En concreto, las leyes 1266 de 2008 y 1581 de 2012, bajo el cumplimiento de las cuales ya es posible procesar la información que los consumidores financieros autoricen expresamente.

A diferencia de las regulaciones en lugares como México, Brasil, Reino Unido y la Unión Europea, en el caso de Colombia este decreto no contempla que la apertura de datos financieros mediante las API bajo un modelo de *open banking* sea de carácter obligatorio en el país.

El decreto tampoco menciona cuáles tecnologías o mecanismos serán los preferidos para acceder a los datos de los consumidores por parte de terceros, abriendo la puerta a agregadores y Fintech.

En el caso particular de este proyecto, el agregador o Fintech (definido en la regulación como “corresponsal”), deberá alinearse a lo que se establece en el *artículo 21, literal e) de la Ley Estatutaria 1581 de 2012 y el Decreto 1074 de 2015*: “e) Impartir instrucciones sobre las medidas y procedimientos necesarios para la adecuación de las operaciones de los responsables y encargados del tratamiento a las disposiciones previstas en la presente ley”.

“La implementación del *open banking* es un cambio de paradigma en la educación sobre la propiedad y el buen uso de los datos de las personas, mientras la regulación colombiana protege la intimidad de los usuarios bancarios” (*La República*, 2022).

El tema de los datos personales y financieros es muy sensible en el ordenamiento jurídico colombiano. La intimidad es un derecho fundamental consagrado en el artículo 15 de la Constitución Política del año 1991. Además, en 2008 se expidió la Ley 1266, con el objetivo de proteger únicamente los datos personales comerciales y financieros de las personas naturales y jurídicas.

Adicionalmente, la jurisdicción colombiana ha adoptado un modelo híbrido para la protección de los datos de las personas, ya que sumó a la ley de 2008, expedida exclusivamente para datos comerciales y financieros, la 1581 de 2015, que se encarga de regular al sector real en esta materia. Así, se configura un sistema robusto, respaldado por la primera ley, en el sector financiero, pero con un alcance igual de potente en el manejo de datos fuera de ese sector.

Debido a lo mencionado, será importante seleccionar a un proveedor o agencias/firmas de analítica de datos que cumplan con ISO27001 y con estándares superiores a los marcados por la regulación colombiana para asegurar que se cubra tanto la normativa actual como las mejoras futuras.

11.2. Plan Nacional de Desarrollo 2023-2026

Adicionalmente, el proyecto de Ley Plan Nacional de Desarrollo 2023-2026, que fue revisado por el Congreso, con fecha límite a mayo 2023, (y convertido después en Ley 2294 de 2023), con avances importantes en lineamientos sobre esquemas de trabajo de datos abiertos (artículo 71), así como la portabilidad tanto de productos como información financiera entre las diferentes IF (artículo 75), por lo que se vislumbra un ambiente propicio sobre el uso de datos alternativos para promover la competencia y la innovación en aras de la inclusión financiera y crediticia (Departamento Nacional de Planeación, 2023).

11.3. Normativa local e internacional de manejo de datos personales

Si los datos alternativos se utilizan para una calificación crediticia, pueden estar incluidos en la definición de un informe del consumidor y estar sujetos a las normas y reglamentos de organismos que regulan el manejo de información localmente, como lo serían la Ley 1266 de 2008, también conocida como Ley de Habeas Data, o la Ley 1581 de 2012, en la cual se establece el régimen general de protección de datos personales en Colombia.

12. Entorno normativo

El presente capítulo describe una serie de recomendaciones para las IF, aliados financieros o agencias/firmas de analítica de datos que tienen en sus planes ser usuarios y consumidores de datos alternativos, para que seleccionen y usen de la mejor manera las fuentes de datos alternativos que mejor sirvan a los propósitos de otorgar créditos al segmento de las EPC o micronegocios.

12.1. Tácticas para mejorar la política de crédito o poder predictivo de modelos de crédito

Cómo se indicó en el Capítulo 2 “Contexto sobre las fuentes de datos alternativos”, frecuentemente los solicitantes de crédito con negadas causas, donde el resultado de su análisis crediticio no es lo suficientemente contundente como para aprobarlo o negarlo inmediatamente. Esto se debe por lo general a la falta de información de actividad económica o historial crediticia, por lo cual la solicitud termina siendo negada o desistida pero no por que se haya confirmado una evidente falta de capacidad de pago o un evidente mal comportamiento de pago, simplemente la falta de información produce incertidumbre y un riesgo asociado a él por lo que las IF deciden no otorgar.

Sin embargo, existen escenarios donde un solicitante puede ofrecer documentación o evidencia adicional para subsanar esa incertidumbre. De esa manera, los resultados favorables de un análisis con datos alternativos, como lo puede ser un buen puntaje con un modelo alternativo, pueden ayudar a subsanar esa incertidumbre.

De igual manera, una IF puede ofrecer a los solicitantes que resultaron negados por causales blandas, pero a los cuales se les pudo recolectar datos alternativos relevantes, términos menos favorables, como una tasa de interés más alta o un límite de crédito más bajo.

A continuación, se describen algunas posibles tácticas para que las IF puedan implementar durante un piloto de uso de datos alternativos, cuando se opta por la estrategia de mejorar la política de crédito o el poder predictivo de los modelos de crédito actuales de la IF.

- a. **Cambios en políticas de admisión:** utilizar *insights* construidos con datos alternativos para ser considerados en la política de crédito. Por ejemplo, establecer un tratamiento particular para micronegocios que tengan mal desempeño en pago de impuestos, califiquen bajo en “honestidad” en una evaluación psicométrica o presenten algún comportamiento específico en su servicio de suscripción de Netflix obtenido de un agregador de finanzas abiertas.
- b. **Agregar instancias de análisis al proceso:** utilizar resultados del *score* alternativo para dar tratamiento particular a micronegocios. Por ejemplo, en el evento que existiera una limitación de fondeo para ofrecer crédito se podría priorizar el otorgamiento a solicitantes que al obtener un excelente resultado en su *score* alternativo se esperaría que tuvieran menor probabilidad de incumplimiento y luego entonces representan menor riesgo.
- c. **Enriquecer modelos de *score* de crédito:** utilizar variables (también llamada “variable estadística”) construidas con datos alternativos para combinarlas con variables construidas de fuentes de datos tradicionales, que la IF recolecte para combinar con los modelos actuales o crear nuevos modelos híbridos.
- d. **Mejorar la oferta de producto para mantenerla competitiva:** utilizar los datos adicionales para entender mejor al solicitante y poder ofrecer productos más contextuales o ajustados a las necesidades del segmento. Por ejemplo, ofrecer tasa preferencial a solicitantes aprobados con excelente resultado en su *score* alternativo para disminuir la carga financiera del micronegocio y producir fidelidad con la IF (Rojas, 2014).

12.2. Capacidades técnicas de la organización usuaria de datos alternativos

El impacto que puedan tener las fuentes de datos alternativos dependerá de la posibilidad de poder combinarla con otras tecnologías *fintech*, como por ejemplo IA, *big data* y análisis predictivo (Monsberger y Del Rosal, 2021), que permitan complementar o suplir el conocimiento experto de las IF y ponerla a disposición de múltiples financiadores que no tengan la misma experiencia en evaluación de riesgo crediticio en segmento de poca o nula información en centrales de información.

Las metodologías de *scoring* alternativo también requieren el desarrollo de ciertas capacidades como parte del proceso de adopción, incluidas las siguientes:

- Gestión de datos: la capacidad de recopilar, limpiar y gestionar datos alternativos para la calificación crediticia.
- Automatización de plataforma: la capacidad de desarrollar una plataforma de *software* para lograr la automatización de la suscripción de créditos y el seguimiento de las solicitudes de crédito de las mipymes.
- Innovación de modelos: la capacidad de formular modelos alternativos de calificación crediticia utilizando IA y aprendizaje automático.

Al considerar lo anterior si se desean elaborar modelos de crédito alternativo para el segmento EPC que pueda ser usado por diferentes IF o aliados financieros, se sugiere el involucramiento de una agencia/firma de analítica de datos independiente con experiencia o capacidades probadas en al menos los tres aspectos mencionados anteriormente.

12.3. Definición del nivel de procesamiento de datos de los resultados que entregará el proveedor de datos alternativos

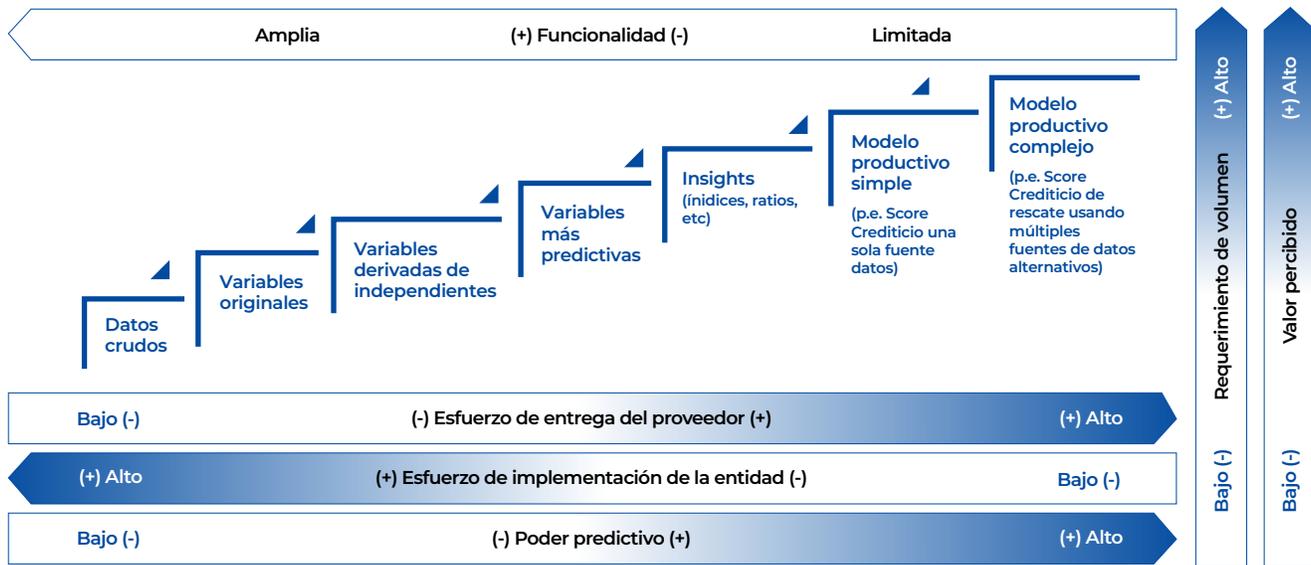
De acuerdo con lo explicado en el capítulo 2, la entidad o aliado financieros deberán definir de acuerdo con el nivel de funcionalidad esperado, el requerimiento de volumen, el valor percibido, poder predictivo y el esfuerzo de implementación esperado.

Si se espera que la entidad o el aliado financieros realicen poco esfuerzo en el desarrollo de un modelo predictivo, entonces se deberá optar por una fuente y proveedor que ofrezca un modelo con un alto poder predictivo.

Las IF o agencias/firmas de analítica de datos pueden usar resultados con menor nivel de procesamiento como los *insights*, así como variables alternativas para desarrollar las estrategias de aumentar niveles de aprobación, mejorar la política o poder predictivo de modelos de crédito actuales y mejorar la eficiencia del proceso o canal de admisión (figura 12-1).

Es por tanto que se recomienda que se trabajen con los proveedores en un inicio solo puntajes, *insights* y las variables más predictivas que tenga el proveedor al momento, acorde con el producto y segmento más afín.

Figura 12-1. Data Journey de datos alternativos



Fuente: elaboración propia de acuerdo con experiencia del equipo consultor.

Lo anterior se debe a que el uso de otros tipos de resultados puede deberse a los siguientes factores:

- **Altos esfuerzos de entrega del proveedor:** producto de contemplar mayor alcance en los contratos con los proveedores, lo que seguramente conllevaría a mayor complejidad en el contrato y proceso de contratación, así como al aumento en costos e inversión.
- **Alto esfuerzo de implementación:** ya sea que se implemente directamente con una institución financiera o con el apoyo de una agencia/firma de analítica de datos; ambas organizaciones requerirán recursos para realizar actividades de ingeniería de datos y analítica, de manera que se requerirán recursos para el fortalecimiento interno principalmente en los equipos de ingeniería de datos, ciencia de datos, modelamiento y riesgo para poder hacer las implementaciones.
- **Bajo poder predictivo:** el uso de fuentes de datos con mayor nivel de procesamiento supone un mayor modelo predictivo. Como, por ejemplo, el uso de un modelo que ya fue entrenado en el producto y segmento objetivo. Por el otro lado, los resultados que han tenido un bajo nivel de procesamiento, como los datos crudos, generalmente suponen poco o nulo poder predictivo, a menos que el proveedor o recursos internos en la organización tengan la experiencia probada en el uso de estos en el producto y segmento objetivo.
- **Alto requerimiento de volumen:** ciertos resultados, como por ejemplo un puntaje de score alternativo que use múltiples fuentes, puede tener un muy buen poder predictivo, sin embargo, para desarrollar o entrenar el modelo de esta complejidad que produzca dicho resultado seguramente requerirá de mucho volumen de datos, tanto de solicitudes de crédito, recolección de datos alternativos como datos de repago de los créditos otorgados. Esto supone unas inversiones relevantes, debido a los recursos necesarios de parte de los proveedores, así como internos de las organizaciones y los recursos adicionales de VaR, que conllevan a las posibles pérdidas adicionales al probar con un número mayor de individuos en la muestra de entrenamiento del modelo.

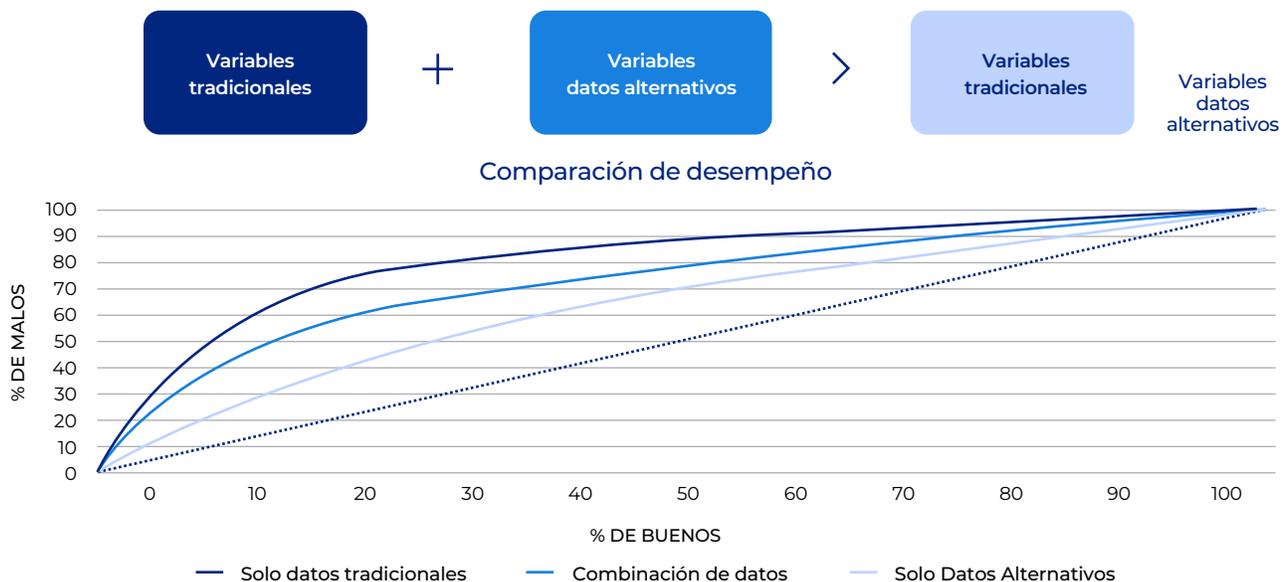
- **Menor valor percibido:** como consecuencia de los puntos anteriores, resultados de datos alternativos que tengan poco nivel de procesamiento, como los datos crudos o variables originales, si bien la funcionalidad para probarlos y del proveedor para entregarlos pueden ser muy convenientes, el valor percibido a las estrategias mencionadas pueden ser pocas debido principalmente a que requerirán mucho ejercicios de procesamiento interno, desarrollo de múltiples casos de uso para implementarlos, ejercicios de prueba y error de todo lo anterior debido a su poco poder predictivo.

12.4. Combinación de fuentes de datos

Los datos alternativos agregan valor predictivo a los modelos de riesgo crediticio basados en datos tradicionales. Algunas consideraciones para tener en cuenta en el impacto de combinar fuentes de datos son: (a) el poder predictivo de los modelos existentes (tanto el tradicional como el alternativo), (b) la solidez de la relación del cliente con la IF, (c) ortogonalidad de los datos.

La figura 12-2 expone el resultado de un proyecto que FICO realizó para una cartera de préstamos personales (FICO, 2022). Las variables de crédito tradicionales capturaron más valor que las de datos alternativos y se identificó un alto grado de superposición entre los dos. Sin embargo, al combinar las variables de los datos tradicionales y alternativos (y comprender la superposición para no ponderar en exceso la contribución de ciertas variables), se produjo un modelo más poderoso.

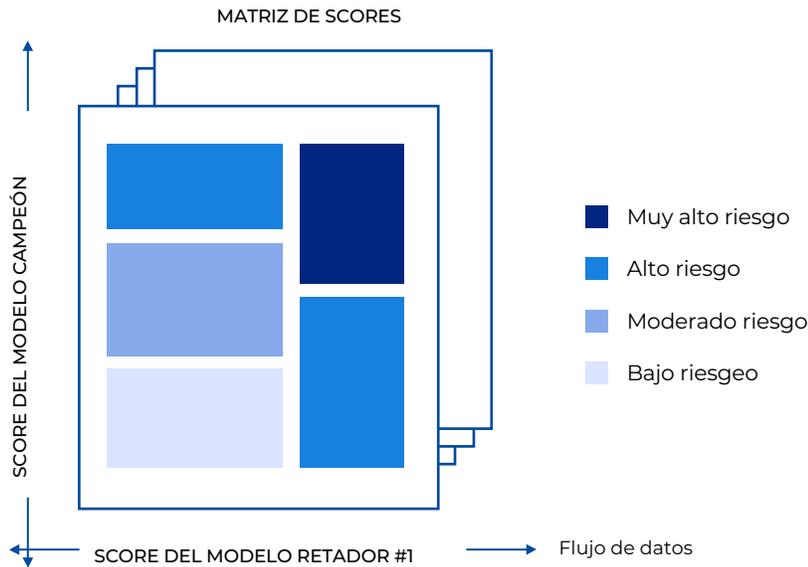
Figura 12-2. Comparación de desempeño en combinación de fuentes de datos



Fuente: FICO blog.

Los modelos de calificación crediticia convencionales y alternativos pueden coexistir. Se identificaron estudios en los cuales la combinación de fuentes incrementa el poder predictivo de modelos y *score cards* de entidades financieras. La IF o deberá definir e implementar técnicas para combinar ambas calificaciones, como el enfoque campeón-retador y matrices de *scores* para que la gestión del riesgo de crédito tome las decisiones de política de crédito (Hong Kong Applied Science and Technology Research Institute, 2020) (figura 12-3).

Figura 12-3. Técnicas de combinación y uso de fuentes de datos. Campeón retador y matrices de puntajes

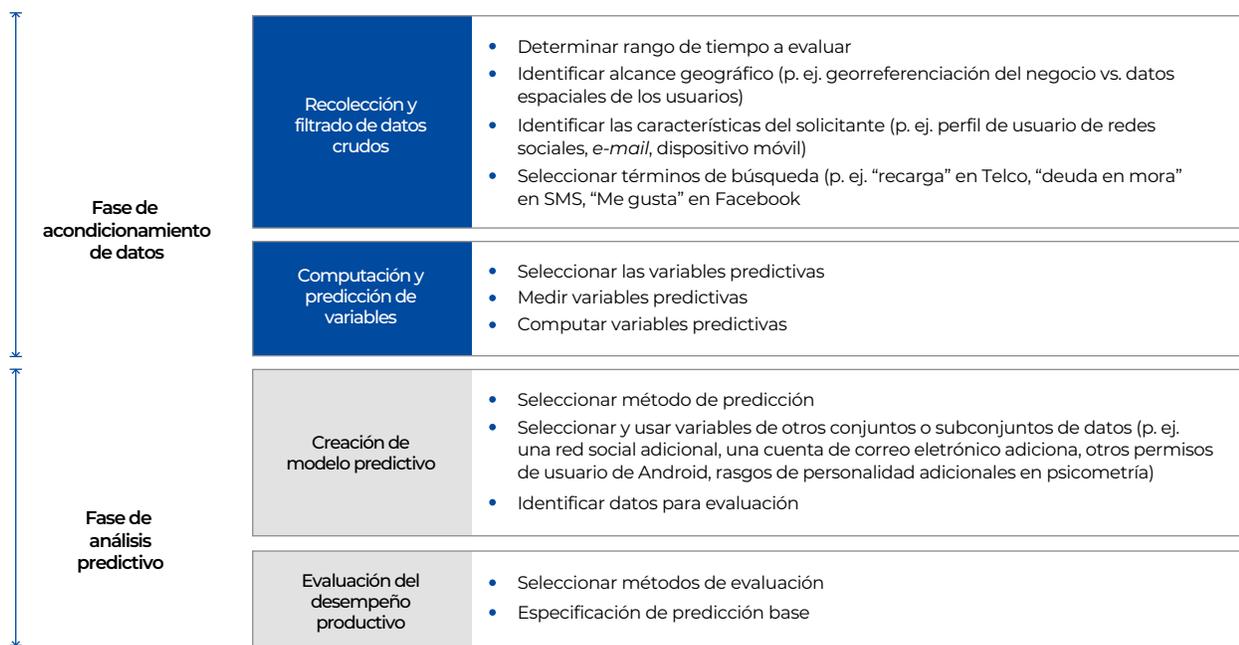


Fuente:FICO

12.5. Esquema de trabajo propuesto para el desarrollo de modelos con datos alternativos

Alineado con las modalidades antes descritas (Putra S. G., 2018) propone un esquema para la evaluación de datos alternativos con enfoque a micronegocios (figura 12-4).

Figura 12-4. Esquema de trabajo para análisis predictivo con datos alternativos



Fuente: elaboración propia con información de (Putra S. G., 2018).

El esquema de trabajo se compone de cuatro etapas (figura 12-4). Las dos primeras, se denominan “fase de acondicionamiento de datos”, mientras que las dos últimas se denominan “fase de análisis predictivo”. La primera etapa es la recopilación y el filtrado de datos sin procesar, que incluye determinar la plataforma como fuente de datos, ventana de tiempo y término de búsqueda para obtener la información deseada. La segunda etapa es calcular la variable (también llamada “variable estadística”) predictora, que incluye seleccionar métricas de datos alternativos como variables predictoras y calcular las variables deseadas a partir de datos crudos de fuentes alternativas.

En un proyecto que FICO realizó para una cartera de origen de préstamos personales, las características de crédito tradicionales capturaron más valor que las de datos alternativos (con los datos alternativos capturando alrededor del 60 % del poder predictivo), y hubo un alto grado de superposición entre los dos. Sin embargo, al combinar las características de los datos tradicionales y alternativos (y comprender la superposición para no ponderar en exceso la contribución de ciertas variables), se pudo producir un modelo más poderoso (FICO, 2022).

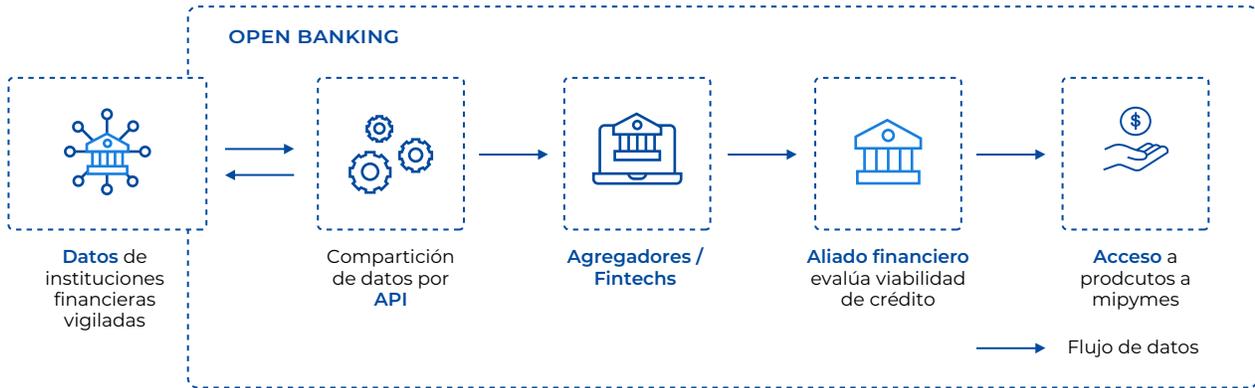
En psicometría concretamente, un estudio realizado en Indonesia tuvo como objetivo investigar si un sistema de calificación demográfico y psicométrico produciría un resultado más preciso en la medición del riesgo crediticio del propietario de las mipymes, en comparación con el sistema de puntuación crediticia con solo factores demográficos (Artarti y Shahnaz, 2018). Esta investigación mostró que el uso de un modelo combinado de datos demográficos y psicométricos tuvo mayor poder predictivo en la probabilidad del riesgo crediticio del prestatario⁴⁰; lo que concluye que la personalidad del prestatario lleva a la evaluación de su riesgo más completo. La combinación de fuentes de datos en este proyecto aumentará la confianza en la capacidad del prestatario para pagar el préstamo.

12.6. Esquema de trabajo propuesto para el uso de datos alternativos considerando modelos de *open banking* existentes

El modelo que se plantea emplear para este proyecto es uno basado en el modelo 1 de *open banking* (sección 6.2.1), donde se buscará un agregador de datos conectado con distintas instituciones financieras en Colombia, que obtiene datos de usuarios para reforzar las capacidades de *scoring* alternativo, ya sea para las IF, así como para actores de la banca pública de desarrollo (figura 12-5).

⁴⁰ El AUC en el modelo combinado tuvo 0,80 puntos y el modelo demográfico 0,78 puntos. Se tuvo una ganancia de 0,02 puntos absolutos.

Figura 12-5. Ilustrativo del modelo planteado, en que, a través de las API de banca, se interactúa con la información financiera de usuarios bancarios mediante un agregador



Fuente: elaboración propia.

Sin embargo, destacando las capacidades y estado de los actores actualmente también se plantea un modelo similar, en el que a través de una alianza con una entidad que ya dispone de canales digitales, se pueda hacer uso del modelo basado en *open banking* (figura 12-6).

Figura 12-6. Modelos planteados para emitir créditos basados en las fuentes priorizadas

	MODELO 1A: Aliado financiero directamente	MODELO 1B: Marketplace + firma de analítica + aliado financiero
Productos de cara al cliente	Aliado financiero (IMF, Fintech, banco, etc.)	Capacidades del marketplace o generador de leads
Onboarding y KYC		
Compartición de datos e integraciones	Capacidades del aliado financiero	
Análítica Infraestructura Gobernanza de datos	Capacidades de la firma de analítica	
Análisis de créditos y detección de fraude	Capacidades de aliado financiero	Capacidades de la firma de analítica
Emisión de crédito		Capacidades de la banca de desarrollo
Provisión de crédito (quién asume riesgo)	Capacidades de aliado financiero	Capacidades de aliado financiero
Liquidación, disputas, aclaraciones reconciliaciones	Capacidades de aliado financiero	Capacidades de aliado financiero

Fuente: elaboración propia con información derivada del análisis.

12.7. Aplicar principios de gestión de proyectos y de sus restricciones al realizar pilotos o pruebas de concepto

El objetivo de probar un modelo alternativo de crédito es mostrar los resultados que se pueden lograr dentro de las limitaciones que se establecen para cualquiera de ellos. Se recomienda gobernar con especial atención la triple restricción que representa el tiempo, alcance y costo (Esterkin, 2008); donde adicionalmente, por tratarse de un proyecto con alto componente estadístico se sugiere establecer la limitación de “tamaño de muestra”. A continuación, se describen algunas consideraciones para cada una de las cuatro restricciones que delimitarán la ejecución del piloto o prueba de concepto sugerido.

12.7.1. Alcance

La implementación de datos alternativos para garantizar un rendimiento positivo deberá contar con pruebas de concepto o pilotos cuidadosos. Existen tres pasos para una prueba exitosa:

- a. Validación de la población actual.
- b. Tratamiento de valores faltantes.
- c. Recalibración del modelo con puntuación retroactiva para evaluar las mejoras en el rendimiento (Urjanet, 2019).

El uso de metodologías ágiles de gestión del proyecto para este tipo de despliegues puede hacer muy eficiente el objetivo de un piloto o PoC. Tener una definición de primeras entregas parciales bajo un cronograma de hitos que brinden seguridad a la organización de los logros y aprendizajes que se tienen en una PoC apoya a que se movilicen los recursos en la evolución del proyecto (Dietrich, De Souza, y Guerreiro, 2020). Se sugiere limitar las actividades a las únicamente vitales de acuerdo con las métricas de éxito del piloto o PoC. Por ejemplo, integraciones o automatizaciones.

Existen múltiples consideraciones a tener en cuenta en el alcance al momento de evaluar y seleccionar fuentes y proveedores de datos alternativos, sin embargo, se sugiere estructurar RFI o RFP principalmente sobre estos (tabla 12-1).

Tabla 12-1. Consideraciones y ejemplos para tener en cuenta en el alcance del proyecto

Consideraciones en el alcance	Ejemplos
Cantidad de fuentes de datos	<ul style="list-style-type: none"> • <i>Open banking</i>. • Psicometría. • Transacciones públicas. • Otras fuentes no priorizadas.
Cantidad de subconjuntos de datos	<ul style="list-style-type: none"> • Datos de monederos electrónicos datos de personalidad, datos fiscales, etc. • Puntajes de crédito vs. variables.
Cantidad de despliegues	<ul style="list-style-type: none"> • Cantidad de productos financieros. • Cantidad de flujos de proceso de crédito. • Cantidad de aliados financieros participantes en el piloto. • Cantidad de modelos de <i>score</i> a usar.
Cantidad de transacciones	<ul style="list-style-type: none"> • Mil solicitudes de crédito al mes. • Hasta cinco consultas de información por trimestre.

Tabla 12-1. Consideraciones y ejemplos para tener en cuenta en el alcance del proyecto (continuación)

Consideraciones en el alcance	Ejemplos
Complejidad de la instalación inicial	<ul style="list-style-type: none"> Integración “sin código” vs. “completa” con soporte técnico. Integración de <i>end points</i> (puntos finales) de agregación o <i>end points</i> de enriquecimiento. Obtención de resultados en tiempo real a través de API. Obtención de resultados asincrónicos o en <i>batch</i>, vía FTPS (o FTP sobre SSL, glosario de términos) o sincronización de repositorio S3*.
Servicios adicionales	<ul style="list-style-type: none"> Construcción o entrenamiento de un modelo. Realización de revisiones de desempeño del modelo. Consultoría de enriquecimiento o construcción de variables usando datos alternativos.

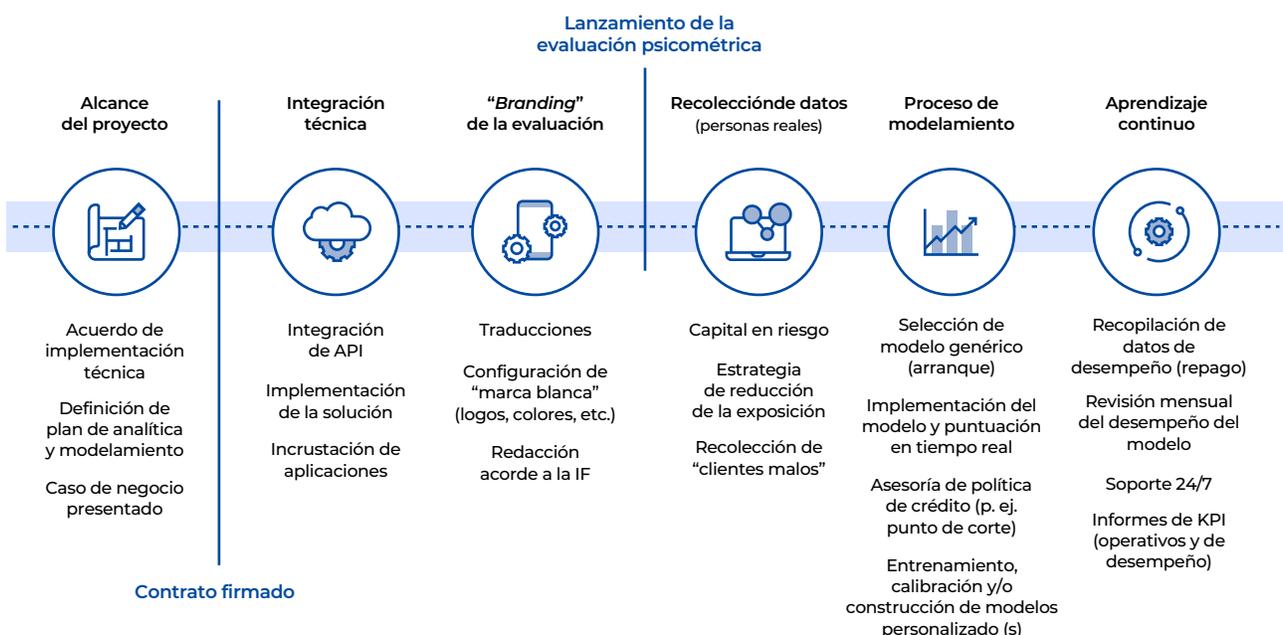
* Simple Storage Service (S3) es un servicio de almacenamiento de objetos en la nube pública de la empresa Amazon Web Services.
Fuente: elaboración propia.

12.7.2. Tiempo (periodos de evaluación)

Con el tiempo, una vez que la cartera haya madurado, muestra si el uso de datos alternativos le permitieron aprobar más solicitantes dentro de sus niveles predeterminados objetivo, controlando por ciclo comercial. Pero si está trabajando con préstamos de 24 a 36 meses, pueden pasar tres o cuatro años antes de que pueda evaluar completamente el impacto del uso de datos alternativos, momento en el cual es posible que la capacidad de atención interna ya se haya agotado.

Se recomiendan utilizar las técnicas de análisis de *early defaults* (mora temprana), para tomar decisiones a lo largo de este periodo de evaluación, en particular si se están utilizando diferentes fuentes de datos o técnicas, como las ya mencionadas en alcance (figura 12-7).

Figura 12-7. Hitos de un lanzamiento de solicitud de crédito con psicometría



Fuente: LenddoEFL

12.7.3. Tamaños de muestra

No hay un número fijo de créditos “malos” que se requieren para construir un modelo de riesgo de crédito estadísticamente aceptable, ya que esto depende de varios factores, como el tamaño y la diversidad de la cartera de préstamos, el nivel de precisión deseado y la complejidad del modelo. En general, una muestra más grande de créditos “malos”, de acuerdo con la definición del umbral conducirá a un modelo de riesgo crediticio más preciso y sólido. Sin embargo, no existe un número específico que pueda considerarse universalmente aplicable u óptimo. Algunos expertos de la industria sugieren que puede ser necesario un mínimo de 100-200 préstamos incobrables para construir un modelo confiable de riesgo crediticio, mientras que otros recomiendan un tamaño de muestra de al menos quinientos o más (Crouhy, Galai, y Mark, 2006).

Con base en lo anterior se recomienda que para entrenar un modelo se tengan datos de al menos quinientos créditos “malos”, de un total de no menos del doble de créditos “buenos”, es decir, que habrán de otorgarse no menos de 1.500 créditos en la etapa de piloto y que se tengan no menos de quinientos para construir un modelo con una o varias fuentes de datos alternativos.

Vale la pena tener en cuenta que la definición de tamaños de muestra muy seguramente incidirá en el tiempo o duración del proyecto; es decir, si se desean obtener seiscientos créditos malos para entrenar un modelo seguramente se requerirá un tiempo adicional para llegar a esa cantidad de créditos, de los que se requeriría si solo se definen quinientos.

12.7.4. Costo o presupuesto

El costo directo de las soluciones de fuentes de datos alternativos es uno de los principales rubros dentro del presupuesto de un proyecto de esta naturaleza y va directamente relacionado con las cuatro restricciones mencionadas anteriormente, es decir, si se considera un alcance mayor y un cronograma más extenso seguramente se tendrán mayores costos por licencias u honorarios en la implementación.

En la tabla 12-2, se describen los principales detonadores de costo y las recomendaciones para el segmento de EPC en Colombia en el marco de un proyecto piloto.

Tabla 12-2. Detonadores de costo y recomendaciones para tener en cuenta en un proyecto piloto

Detonadores de costo	Recomendación para el segmento EPC Colombia
Selección de fuentes de datos	<ul style="list-style-type: none">• Usar <i>open banking</i>, psicometría y transacciones públicas únicamente.
Subconjuntos de datos – <i>Open banking</i>	<ul style="list-style-type: none">• Datos crudos de Nequi y Daviplata.
Subconjuntos de datos – psicometría	<ul style="list-style-type: none">• Resultados procesados de <i>score</i> psicométrico.• Resultados semiprocesados de <i>insights</i> de personalidad.
Subconjuntos de datos – Transacciones públicas	<ul style="list-style-type: none">• Datos crudos de registro social.• Datos crudos de fuentes de vulnerabilidad de población.• Datos de servicios públicos.

Tabla 12-2. Detonadores de costo y recomendaciones para tener en cuenta en un proyecto piloto (continuación)

Detonadores de costo	Recomendación para el segmento EPC Colombia
Cantidad de transacciones a contratar	<ul style="list-style-type: none"> Considerar todo el flujo de proceso y calcular el volumen de transacciones a contratar teniendo en cuenta el principio del “embudo” de este flujo. Por ejemplo, si se estima que, por cada cincuenta solicitudes de crédito, solamente se desembolsa uno, entonces esta razón se debe tener en cuenta en el volumen por contratar.
Rampa de transacciones	<ul style="list-style-type: none"> Se debe considerar que los volúmenes por lo general son incrementales a lo largo del tiempo y conforme madura el piloto por lo que no se recomienda contratar sobre volúmenes lineales sino volúmenes en rampa incremental y de preferencia en costo por transacción de manera que no existan ineficiencias en el volumen contratado.
Complejidad de la instalación en la primera etapa (prueba de concepto)	<ul style="list-style-type: none"> Se recomienda hacer inicialmente una prueba de concepto o PoC para probar el resultado de los datos de preferencia con poca inversión, corto alcance o nulos requerimientos tecnológicos, ante lo cual un proyecto sin integraciones optimizaría los recursos respecto al tiempo y el presupuesto en esa etapa inicial. Evaluar la opción de hacer un ejercicio de <i>backtesting</i> inicial si el proveedor y el set de datos de prueba lo permiten. Probar con clientes reales de manera asíncrona o por lotes en <i>batch</i> vía FTP (FTP sobre SSL), o sincronización de repositorio S3* Evitar tomar definiciones e inversiones para realizar procesos de otorgamiento en tiempo real. Definir un primer grupo de fuentes, subconjuntos o resultados más relevantes y probar con ellos.
Complejidad de la instalación en la segunda etapa (piloto)	<ul style="list-style-type: none"> Integración de <i>end points</i> de agregación o <i>end points</i> de enriquecimiento básicos. Definir un segundo grupo de fuentes, subconjuntos o resultados más relevantes y pilotear con ellos.

* Simple Storage Service (S3) es un servicio de almacenamiento de objetos en la nube pública de la empresa Amazon Web Services. Fuente: elaboración propia con información derivada del análisis.

Con base en las anteriores recomendaciones, así como las inversiones estimadas de las fuentes de datos priorizadas de los capítulos anteriores, se estimaron los siguientes escenarios de costos para proyectos de implementación de fuentes de datos alternativos en modelos de crédito (tablas 12-3 y 12-3).

Tabla 12-3. Escenarios de costos de instalación inicial en proyecto

	Trial o prueba de concepto	Piloto	Producción
Duración	Entre 1 mes y 3 meses	Entre 4 meses y 18 meses	+ 12 meses
Costo	Entre 1.000 y 5.000 USD	Entre 5.000 a 12.000 USD	Entre 7.000 a 44.000 USD

Fuente: elaboración propia, a partir de investigación previa, bases de datos y estudios de cliente oculto.

Tabla 12-4. Escenarios de costos recurrentes mensuales en proyecto

Volumen de consultas o transacciones mensuales	Trial, prueba de concepto o piloto (USD)	Producción (USD)
1-3.000	Entre 1.000 y 6.000	Entre 4.000 y 000
3.001 – 12.000	Entre 1.600 y 15.000	Entre 4.600 y 15.000
12.001 – 30.000	Entre 4.400 y 20.000	Entre 7.000 y 25.000
30.001 – 50.000	Entre 6.000 y 25.000	Entre 10.000 y 30.000
50.001+	Entre 10.000 y 35.000	Entre 15.000 y 50.000

Fuente: elaboración propia, a partir de investigación previa, bases de datos y estudios de cliente oculto.

12.8. Recomendaciones sobre la disponibilidad de los datos

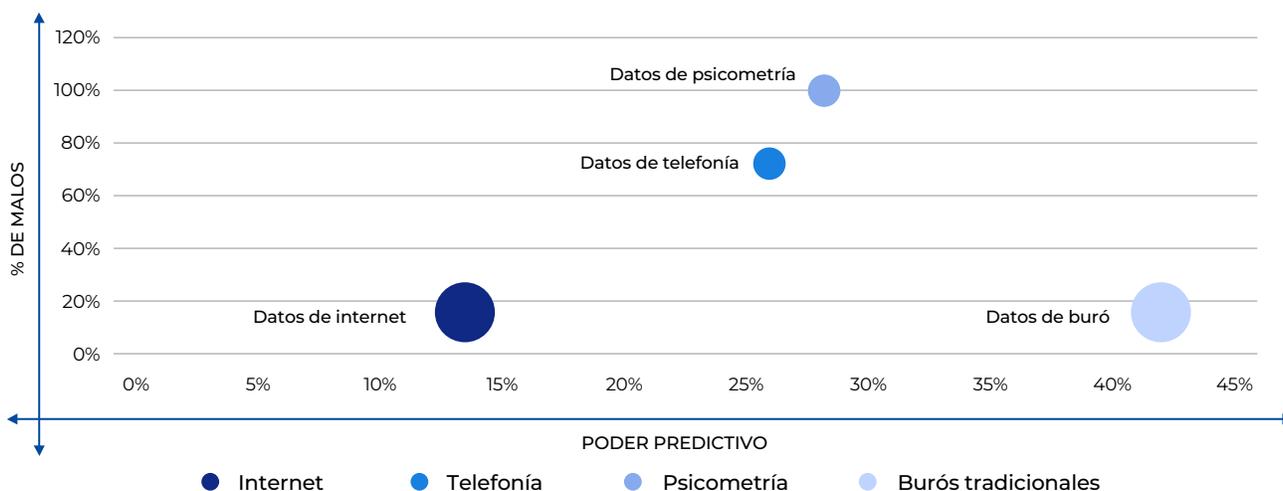
Independientemente del uso de las técnicas, así como las capacidades mencionadas anteriormente, la selección de la fuente o las fuentes de datos deberá considerar las restricciones que existen entre disponibilidad y poder predictivo. Los datos financieros provenientes de buró siguen siendo de las fuentes más predictivas, sin embargo, como se ha mencionado anteriormente tiene una disponibilidad limitada (United Nations Development Programme Pakistan, 2022).

Por otro lado, los datos psicométricos bajo el fundamento de que se obtienen a partir de la personalidad y todos los individuos tienen una personalidad, entonces tienen una disponibilidad muy cercana a cien por ciento, como se mencionó en el Capítulo 5 referente a los “Datos psicométricos”. De la misma manera en cobertura le siguen los datos de dispositivos móviles cuya cobertura está creciendo constantemente; sin embargo, ambas presentan un reto de poder predictivo.

De acuerdo con lo que se mencionó en el Capítulo 6 “Datos de *open banking*”, Nequi y Daviplata han abierto sus datos a través de API y han implementado alianzas con distintos agregadores de datos, entre los que se destacan los ya mencionados Belvo y Finerio Connect, por lo que resulta muy interesante para este proyecto incluir en el análisis soluciones que tengan estas alianzas que permitirían aumentar la disponibilidad y cobertura de datos para el segmento de las EPC.

Al usar una nueva fuente de datos se plantean varios desafíos, se recomienda que los resultados de las fuentes de datos que se usen ya sean variables, *insights* o modelos estén en capacidad de poder realizar *backtesting* (o pruebas retrospectivas), de manera que se tenga la capacidad de evaluar la precisión del modelo genérico, por ejemplo por sobreajuste o un ajuste insuficiente, previendo de esta manera situaciones donde el modelo genérico funcionaba bien en los datos de entrenamiento pero mal en los datos de prueba o viceversa, lo que llevaría a predicciones inexactas y resultados poco confiables (figura 12-8).

Figura 12-8. Balance disponibilidad vs. poder predictivo en las fuentes de datos



El tamaño de la burbuja se aproxima a la facilidad de implementación. Burbuja más grande = Menores costos

* Coeficiente de Gini fuera de la muestra con modelo y muestras comparables.

** Tasa de aciertos en la muestra de solicitantes de bajos ingresos en mercados emergentes.

Fuente: tomado del Development Advocate Pakistan. Inclusive Growth in times of uncertainty.

12.9. Recomendaciones sobre sesgos y prácticas de discriminación

Si bien algunas fuentes de datos alternativos están en posibilidad de obtener miles o hasta centenas de miles de puntos de datos de un solicitante y algunos proveedores tienen amplias capacidades analíticas como para desarrollar modelos de IA; estos modelos usados para puntaje de crédito en ocasiones pueden llegar a carecer de transparencia en cómo se recolectan y procesan los datos y de cómo se produce el resultado (The World Bank Group, 2022). Esto puede llevar a que se tengan sesgos con consecuencias discriminatorias que afectan las buenas de inclusión y justo acceso al crédito.

Ante eso resulta vital contemplar dentro del alcance la revisión de los diccionarios de datos para entender cómo están conformadas las variables consideradas en los modelos y el posible impacto de ellas.

12.10. Establecimiento de criterios de éxito en la POC o piloto

- Cuantitativos
 - Resultados operacionales (duraciones, conversiones)
 - Resultados de predicción (AUC, Gini, Risk Split)
 - Cumplimiento de supuestos del caso de negocio (% aprobación, % default, capital en riesgo, tiempos de maduración de cosechas, ROI)
- Cualitativos (aprendizajes blandos)
- Experiencia del cliente
 - NPS
 - Facilidad de implementación

- Cumplimiento a la promesa de valor
- Servicio al cliente/soport
- Capacidades técnicas del proveedor de fuentes de datos

12.11. Recomendaciones sobre el monitoreo del desempeño de los modelos alternativos

Se recomienda que dentro del alcance y plan de analítica con el proveedor se acuerden revisiones periódicas de desempeño del modelo, que considere lo siguiente:

- **Estabilidad de la población:** comprobación de estabilidad de la población para determinar si los valores de las variables son estables a lo largo del tiempo. Se sugiere que los datos se dividan en segmentos de tiempo para realizar análisis como: índice de estabilidad de la población (PSI), distribuciones de variables, análisis univariante, correlaciones de los predictores y ponderaciones de variables, entre otros.
- **Monitoreo de la cartera:** revisar análisis periódicos de número de desembolsos, tasas de impago y matriz de transición, entre otras.
- **Desempeño del modelo:** seleccionar diferentes muestras de población y variables del modelo para garantizar que este continúe funcionando en todos los casos. Se recomienda que se realicen actividades como análisis de riesgo marginal o *Risk Split* (RS) (por sus siglas en inglés) (división del riesgo), desempeño de clasificación del modelo y verificación del poder predictivo de cada variable.

Se recomienda que se incluyan cláusulas que establezcan situaciones en las cuales, si los modelos no parecen funcionar de manera óptima, el equipo del proveedor comenzará a hacer labores mitigantes como recalibrar los modelos o construir nuevos.

Se recomienda que dentro de los planes y de los presupuestos se tenga en cuenta que las recalibraciones deberían ocurrir con mayor frecuencia al inicio del despliegue de los modelos, por lo cual es importante considerar este soporte técnico analítico dentro de la implementación de las soluciones.

Algunas fuentes de datos alternativos están sujetas a cambios y actualizaciones tanto técnicas como normativas. Por ejemplo, algún subconjunto de datos de *open banking* puede contener información distinta, ya que la API que extrae los datos tuvo alguna actualización; un cambio en el método de definición de un parámetro que podría afectar la categoría de un solicitante en una lista poblacional del estado o algún cambio en las políticas de privacidad de Google que ya no permite recolectar datos de georreferenciación o de movilidad de dispositivos, por mencionar algunos ejemplos. Es, por tanto, que se recomienda considerar dentro del alcance la revisión periódica de los diccionarios de datos para entender cambios en los subconjuntos de datos que pudieran afectar el cómputo de las variables e *insights* que están consideradas en los modelos, así como en la agregación y enriquecimiento de datos que produce el proveedor.

12.12. Incluir documentación técnica del modelo de crédito alternativo en alcance y plan de analítica

Durante el proceso de despliegue y activación se deberá preparar y validar la documentación del modelo. De manera que se pueda validar la lógica con la que el modelo produce la puntuación. Esta documentación deberá contener: información sobre las variables utilizadas en el modelo, los valores admisibles y cómo crearlos y probarlos, especificaciones del modelo y datos para la comprobación de la implementación de este.

Esta documentación es crucial para presentar ante organismos de vigilancia como la SFC, así como para justificar inversiones, fondeos o pilotos de implementación en segmentos o poblaciones similares ante inversionistas, juntas, contraloría o fondos de garantías entre otros.

12.13. Sobre el uso de datos alternativos en EPC en zonas rurales

La geografía donde se ubica la población a ser evaluada mediante el uso de datos alternativos incide en la calidad de estos en su disponibilidad, cobertura, especificidad y ortogonalidad. Si bien el alcance de este estudio se enfocó a geografías urbanas, he aquí algunas consideraciones para el uso en zonas rurales:

Datos de Telco y CDR: puede haber menos datos de telecomunicaciones o datos de “registro de detalle de llamadas” en las zonas rurales, debido a la menor cobertura e infraestructura para las redes móviles en esas regiones. Esto significa que menos personas tienen acceso a teléfonos móviles o señales confiables y, por lo tanto, hacen menos llamadas o usan menos servicios de datos. Otra posible consideración es que las poblaciones rurales tienden a ser más pequeñas y más dispersas que las urbanas, y pueden tener diferentes patrones y preferencias de comunicación. Por ejemplo, pueden confiar más en las interacciones cara a cara u otras formas de comunicación, audios de WhatsApp o inclusive los teléfonos fijos, que no producen datos de telecomunicaciones.

Datos de redes sociales y huella digital: puede haber menos datos de redes sociales producidas en las zonas rurales principalmente, debido a la falta de acceso a servicios de Internet confiables y asequibles. La conectividad a Internet a menudo es limitada o costosa en regiones remotas, lo que puede desanimar a las personas a usar plataformas de redes sociales o compartir su información personal en línea. La menor densidad de población y una mayor cohesión social en las zonas rurales, podría reducir la necesidad o el deseo de comunicación y redes en línea. Las personas en las zonas rurales pueden tener lazos más fuertes con sus comunidades locales y menos interés en llegar a extraños o conocidos distantes a través de las redes sociales.

Datos de open banking: se espera que existan menos datos bancarios abiertos producidos por la población en las zonas rurales principalmente los aún menores índices de inclusión financiera y alfabetización digital que los residentes urbanos. Sin embargo, de acuerdo con lo que se mencionó en el Capítulo 6 “Datos de open banking”, si se espera que se tenga buena cobertura y disponibilidad respecto al uso de información de billeteras móviles en especial Nequi y Daviplata y adicionalmente la cuenta “Ahorro a la mano”⁴¹.

⁴¹ Es una cuenta de ahorros digital de Bancolombia que se puede abrir y manejar desde el celular o la web, sin necesidad de ir a una oficina.

De la misma manera, los dueños de micronegocios rurales, aunque sí tengan cierto nivel de bancarización, probablemente se tenga el reto de que se observe menor conciencia y confianza en los beneficios y riesgos de los datos abiertos, así como la privacidad y la seguridad de los datos, las ofertas personalizadas y la educación financiera. Esto puede reducir su disposición a consentir para compartir sus datos con terceros o acceder a nuevos productos y servicios financieros a través de canales de banca abierta. Por lo que muy seguramente se requerirán esfuerzos adicionales de confianza, educación y adopción al momento de realizar una implementación de uso de datos alternativos con estas poblaciones.

Datos de transacciones públicas: es posible que la información de fuentes de registro social, base maestra y servicios públicos, entre otras no dispongan del mismo nivel de cobertura, especificidad e inclusive limpieza. Sin embargo, de acuerdo con lo referido por Ágata⁴² podría existir información de ruralidad y otras específicas a desarrollo agrícola que podría llegar a suplir esas deficiencias tales como:

Datos producidos por la banca pública de desarrollo: por ejemplo, Banco Agrario y Finagro.

Datos de convenios institucionales con privados como, por ejemplo:

- **AgrodatAi⁴³:** plataforma digital que ofrece información y transfiere conocimiento a los actores de la cadena del agro para la toma de decisiones y conexión comercial.
- **BloomsPal⁴⁴:** comunidad digital para el abastecimiento y la adquisición eficiente de productos agrícolas, en la cual se conectan compradores y proveedores de la cadena del agro. Dispone de mercado en línea, red de negocios, procesamiento de pagos y logística.
- **Agrapp.co⁴⁵:** plataforma digital que impulsa el sector agrícola mediante acceso a recursos financieros, estructuración, asistencia, comercialización y tecnificación de los cultivos.

13. Recomendaciones sobre capacidades y acciones de la banca pública de desarrollo

La banca de desarrollo es dueña de estrategia de inclusión financiera de la economía popular, por lo cual debe actuar como promotora y garante del desarrollo de iniciativas de inclusión en este segmento, comenzando con la formulación de los proyectos y el diseño de las pruebas y pilotos que impactarán la economía popular con crédito de bajo monto, en ese sentido, se considera que esta banca atiende las siguientes recomendaciones, que permiten alianzas para maximizar sus capacidades y realizar acciones que extiendan el éxito de la estrategia de inclusión en las economías populares.

⁴² Obtenido de reuniones y entrevistas con equipo de Ágata.

⁴³ <https://agrodatai.com>

⁴⁴ <https://www.bloomspal.com>

⁴⁵ <https://www.agrapp.co>

13.1. Habilitar el uso de metodologías de modelamiento usando datos alternativos

Impulsar la regulación en sus niveles más esenciales y operacionales para incentivar a la banca pública de desarrollo, que use modelos de *scoring* alternativo en los procesos de crédito de sus instituciones y que este ejercicio sea liderado por la agencia/firma analítica de datos, propuesta en los esquemas de trabajo mencionados en la presente investigación.

13.2. Vinculación con agencias/firmas de analítica de datos y sus capacidades técnicas

La banca pública de desarrollo puede apalancarse en lo estipulado en el artículo 72 del PND con respecto a sistemas de información estadísticos, en los cuales el Departamento Administrativo Nacional de Estadística (DANE) liderará la iniciativa de sistemas de información enfocados en la economía popular por primera vez con énfasis en fuentes alternativas.

En la tabla 13-1, se recoge una lista corta de posibles agencias/firmas analíticas con las que se podría colaborar.

Tabla 13-1. Posibles agencias/firmas analíticas

Compañía	HQ	Categoría	Sitio web	Año de fundación	Experiencia Colombia	Comentarios
Ágata Data	Colombia	Agregación y analítica de datos	agatadata.com	2021	Sí	Agregador de datos con convenios de múltiples transacciones públicas en Bogotá principalmente.
Teradata	USA	Analítica de datos	teradata.com	1969	Sí	Soluciones robustas y oficina en Colombia
Sinnetic	Colombia	Analítica de datos	sinnetic.com	2009	Sí	Agencia de datos con soluciones <i>ad-hoc</i> .
DataKnow	Colombia	Analítica de datos	dataknow.io	2018	Sí	Agencia de datos con soluciones <i>ad-hoc</i> .
Procalidad Analytics	Colombia	Analítica de datos	procalidad.com	1989	Sí	Produce modelos analíticos enfocados en la transformación digital de las empresas.
Protiviti	Colombia	Analítica de datos	protiviti.com	2002	Sí	Agencia de datos con soluciones <i>ad-hoc</i> .
Seidor Analytics	USA	Analítica de datos	seidoranalytics.com	2002	Sí	Agencia de datos internacional con soluciones <i>ad-hoc</i> .
LiSim	Colombia	Analítica de datos	lisim.com	1996	Sí	Especializada en <i>scoring</i> .
AIS Group	España	Analítica de datos	ais-int.com	1987	Sí	Gestión integral del riesgo de crédito, desde scorings y ratings, sistemas expertos hasta detección de fraude.
NTT Data	Japón	Analítica de datos	co.nttdata.com	1967 (1988)	Sí	Dentro de los 10 principales proveedores de servicios de TI y analítica del mundo. Diseñan, implementan y operan plataformas y aplicaciones bancarias en una amplia gama de servicios y soluciones digitales.

Fuente: elaboración propia, basada en investigación previa de identificación localización de proveduría regional.

13.3. Aproximaciones regulatorias en las primeras implementaciones de modelos con datos alternativos

Cómo ya se mencionó desde una perspectiva regulatoria, los proveedores de fuentes de datos alternativos generalmente no caen bajo los marcos regulatorios existentes, y sus actividades han comenzado a atraer la atención de las autoridades reguladoras.

Sin embargo, de acuerdo con lo establecido en los “principios clave para la efectiva regulación y supervisión de los “proveedores de servicios de reporte de crédito”, la regulación de las nuevas tecnologías requiere de un equilibrio-enfoque que promueve la innovación mientras supervisa sus implicaciones de riesgo (The World Bank Group, 2022). Ante esto se sugiere una aproximación evolutiva, como está contenida en la figura 13-1.

Figura 13-1. Aproximación evolutiva de las implicaciones de riesgo



Fuente:

- a. **Observación y monitoreo de innovaciones (sin intervención):** rol pasivo inicial produciendo mecanismos de observación y monitoreo de lo que implican las innovaciones en el uso de datos alternativos. Las intervenciones serán a demanda conforme al cumplimiento de “criterios de éxito” en una PC.
- b. **Enfoque de supervisión ligero:** en este enfoque las entidades puedan hacer un piloto generalmente posterior a una PoC, de manera que se permita que las entidades hagan un lanzamiento de una solución de recolección y procesamiento de datos alternativos con clientes reales donde se pueda establecer un marco de trabajo más estricto si de considerarse necesario, por ejemplo, en el evento que los “criterios de éxito” del piloto no se estén cumpliendo.
- c. **Promoción con facilitadores de innovación:** los *hubs* de innovación o el *sandbox* regulatorio de Colombia ofrece oportunidades para usar datos alternativos en los modelos de riesgo de crédito al ser una herramienta que busca promover la innovación financiera para mejorar y ampliar la oferta de servicios facilitando el acceso de los colombianos a estos productos de manera eficiente, ágil y segura (Universidad Externado de Colombia, s.f.), lo cual permite a las empresas innovadoras operar temporalmente bajo ciertas reglas que limitan el número de usuarios o el periodo de tiempo en que se puede ofrecer sus productos.

13.4. Promoción con facilitadores de innovación como el *Sandbox* regulatorio

Como se indicó anteriormente la promoción de las metodologías de modelos de crédito con datos alternativos con *hubs* de innovación y *Sandbox* regulatorio supondría la posibilidad de probar nuevas soluciones basadas en datos alternativos como los de *open banking* o psicometría; estos también presentan algunas dificultades tanto en la teoría como en la práctica.

En la teoría establece mecanismos ágiles de transición al proceso de licenciamiento de una entidad financiera regulada (Martínez y Torres, 2020), es decir, aparentemente está orientado a organizaciones que pretenden obtener una licencia de la SFC, pero no es tan clara la contribución hacia los proveedores de la tecnología únicamente, para este caso los proveedores de la tecnología sobre el uso de datos alternativos.

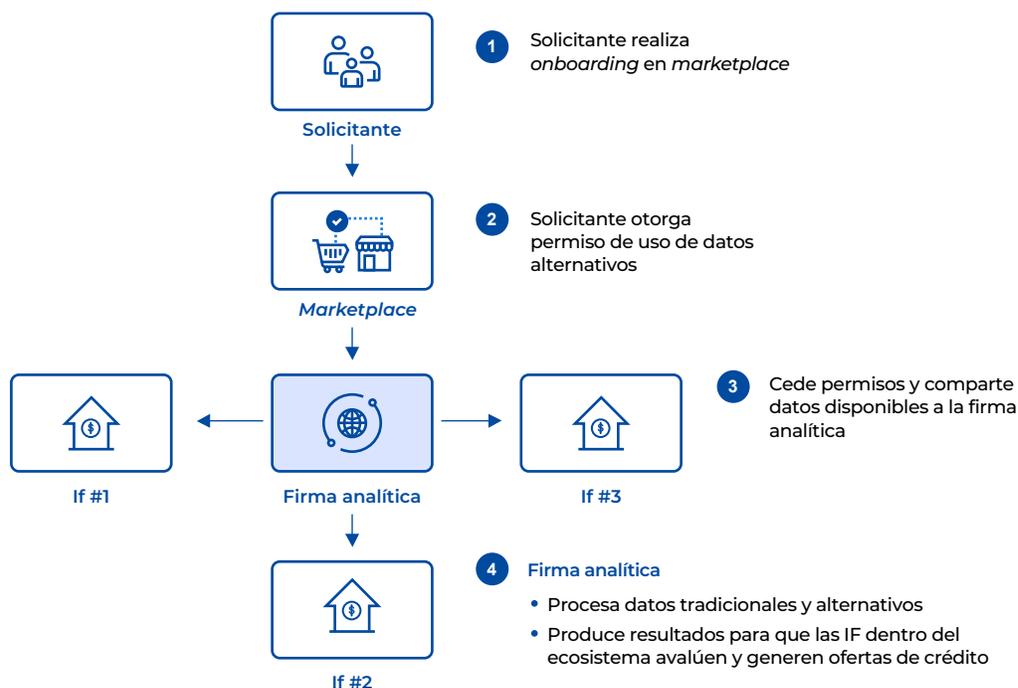
En la práctica, se han desarrollado hasta febrero de 2023, dieciocho iniciativas, que en promedio han tomado catorce meses en desarrollarse y que han derivado en escasos instrumentos normativos (Rincón, 2023); nada relativo a modelos de puntaje de crédito o al uso de datos alternativos en ellos. Por lo que es importante tener en cuenta que el *sandbox* es una herramienta relativamente nueva y que su implementación y uso pueden llevar tiempo y requerir ajustes y mejoras para adaptarse mejor a las necesidades de los *fintech* y del ecosistema financiero en general.

13.5. Adopción de estrategias de *marketplace* para la democratización de los datos para maximizar el impacto con múltiples ofertas de crédito

Tal y como se mencionó en la sección de *open banking*, en cuanto a los modelos de negocio posibles, así como las experiencias que se han tenido en el uso de agencias o firmas analíticas para acompañar el proceso de gestión de los datos alternativos, un punto importante a señalar se debe a que parte del éxito de estas estrategias es la implementación de plataformas estilo *marketplace*, que conecta las necesidades de oferta y demanda.

De esta manera, una solicitud de crédito que llegue a la plataforma, puede ser evaluada con los diferentes datos alternativos y los resultados de este análisis de crédito con modelos alternativos pueden ser entregados a la red de aliados financieros dentro del *marketplace*, de manera que un solicitante pueda ver en acción las ventajas que obtiene de dar permiso de usar sus datos al obtener varias ofertas de crédito; lo cual supone una estrategia de democratización de los datos del solicitante y lleva a una sana y libre competencia entre las IF afiliadas a ese *marketplace* o plataforma, como se plantea en la figura 13-2.

Figura 13-2. Flujo de proceso en la democratización de datos y elaboración de múltiples ofertas de crédito



Fuente: elaboración propia.

En Colombia ya hay experiencias en esta materia; tal y como se mencionó en el apartado referente a la experiencia combinando diferentes fuentes de datos públicos en Bogotá con la participación de la Agencia Analítica Ágata y la SDDE; esta iniciativa corre sobre una plataforma denominada “Crédito Peso a Peso”, en la cual los comerciantes tienen la posibilidad de acceder a créditos de aliados financieros al postularse a través de un formulario⁴⁶. La agencia hace el proceso de calificación alternativa la cual es entregada a las IF aliadas para que ellos usen esos resultados como parte de su proceso de otorgamiento (Secretaría Distrital de Desarrollo Económico, 2023). Si la IF decide aprobar, estas contactarán al comerciante solicitante para ofrecerle opciones de crédito.

Otra experiencia similar se tiene del lado de Bancóldex⁴⁷ con la plataforma Neocrédito⁴⁸, que conecta microempresarios con aliados financieros; la plataforma solicita datos a través de formularios y archivo documental que son trasladadas a IF aliadas para ser evaluadas y en caso positivo, una o varias opciones de crédito pueden ser ofrecidas al solicitante.

Ante esto, la recomendación es intensificar la adopción de estrategias de plataformas *marketplace* alineado al esquema de trabajo propuesto para el uso de datos alternativos considerando modelos de *open banking* existentes (referirse al Capítulo 12. “Recomendaciones de selección y uso de fuentes de datos para el mercado EPC colombiano”), así como a la recomendación de “Vinculación con agencias/firmas de analítica de datos y sus capacidades técnicas” mencionada anteriormente en este capítulo.

⁴⁶ <https://bogota.gov.co/creditopesoapeso/>

⁴⁷ Bancóldex es el banco de desarrollo empresarial de Colombia, que apoya la actividad productiva de las mipymes. <https://www.bancoldex.com/es/sobre-bancoldex/quienes-somos>

⁴⁸ <https://neocredito.bancoldex.com/neocredito/>

De esta manera, la plataforma podría ser un mecanismo de empoderamiento de los solicitantes al ser el eje de la democratización de sus propios datos para obtener opciones de crédito más competitivas. Esto representaría un siguiente paso que le daría escalabilidad a los procesos más comunes de elaboración de nuevos modelos de crédito alternativo que generalmente son desarrollados para una sola IF y podría tener mayor impacto al abordar varias IF a la vez en segmentos comparables. Esto es de particular relevancia considerando que en segmentos con poco o nulo historial crediticio gran parte del costo de adquisición se encuentra en recabar información y si se trata de un crédito de bajo monto esta proporción porcentual del costo de adquisición sobre el retorno de la operación hace que regularmente sea inviable el producto.

13.6. Selección y priorización de IF aliadas para implementar los esquemas de trabajo propuesto

Si bien los primeros ejercicios de estrategias de inclusión mediante el uso de *marketplace* así como agencias/firmas analíticas mencionadas anteriormente han tenido resultados favorables con las IF, es importante priorizar y vincular IF que le den escalabilidad a los programas adoptando los esquemas de trabajo propuestos por la presente investigación donde se recomienda que se tenga una mezcla de diferentes IF de distintos sectores, como bancario, cooperativo, microfinanciero e inclusive *fintech*.

La banca pública deberá apalancarse en los compromisos del Gobierno nacional, en cuanto al fortalecimiento del Grupo Bicentenario como encargado de la estrategia de inclusión en las economías populares (Grupo Bicentenario, 2023); ante lo cual dentro de los primeros pilotos que surjan de acuerdo con lo expuesto en la presente investigación, el Banco Agrario debería ser considerado y priorizado como IF aliada al ser la entidad financiera de primer piso dentro del Grupo Bicentenario.

Adicionalmente considerando que el sector cooperativo es el que tiene más cercanía y penetración en las economías populares la banca pública de desarrollo, deberá priorizar y fortalecer las capacidades mencionadas en las recomendaciones de la presente investigación para priorizar a las principales entidades del sector cooperativo apalancándose en los fundamentos de la estrategia de inclusión de economías populares que describen este sector como eje fundamental (Confederación de Cooperativas de Colombia, CONFECOOP, 2023).

13.7. Diseñar estrategias ante las principales razones de no uso o adopción de fuentes de datos alternativos

Existen diferentes razones por las cuales las IF deciden no iniciar el camino del uso de datos alternativos en sus procesos y modelos de crédito, siendo las mayoría asociadas del desconocimiento sobre estas fuentes de datos, por lo cual, la presente investigación ha considerado estas razones y se han producido secciones de estudio o recomendaciones para mitigarlas, sin embargo se recomienda que la banca pública de desarrollo tenga muy presente estas razones para elaborar estrategias adicionales a las expuestas en esta investigación, que permitan eliminar barreras o impedimentos en la adopción de datos alternativos en los modelos y procesos de crédito por parte de las IF.

Se analizó la literatura con respecto a los principales retos sobre el uso de datos alternativos en general (Exabel, 2022), así como de banca abierta específicamente (SBS, 2022), experiencias registradas de uso de estos datos para amplificación del poder predictivo (Viani B. Djeundje, 2020) y experiencia del equipo consultor.

A continuación, una compilación sobre las principales razones de no adopción de estas fuentes de datos, las cuales la banca pública de desarrollo deberá de considerar:

Desempeño (del modelo)

- Incertidumbre en el desempeño de un modelo de puntuación alternativo.
- Dificultades para justificar el modelo de puntuación alternativo ante un regulador.

Financiero

- Falta de fondos para invertir en un piloto o prueba de un modelo de puntuación alternativo.
- Falta de fondos para poder asumir posibles pérdidas esperadas de un modelo alternativo.

Proceso

- Preocupaciones por la fricción que puede dar al proceso de otorgamiento actual.
- Tiempo de respuesta que puede tener una puntuación.
- Qué el puntaje no tiene suficiente cobertura para calificar el segmento que desean evaluar.

Talento (falta de)

- Recursos necesarios o dedicación del área de tecnología
- Recursos necesarios o dedicación del área de riesgos
- Recursos necesarios o dedicación del área de analítica/inteligencia de negocios.
- Recursos necesarios o dedicación del área de ciencia de datos

“Alfabetización” acerca de datos alternativos

- Desconocimiento de las implicaciones que conlleva la puesta en marcha de una iniciativa de esta naturaleza.
- Desconocimiento acerca de qué fuente de datos es la más apropiada para el segmento al que quieren dirigirse.
- Incertidumbre en costo, tiempo y alcance para construir un caso de negocio sólido.
- Se considera que este tipo de iniciativas son demasiado “disruptivas” y seguramente son difíciles de justificar al interior de la organización.

Canal

- Se considera que no se tiene la madurez suficiente en el canal digital.
- Incertidumbre o falta de claridad debido a demasiadas iniciativas de digitalización de canales en curso.

Dolor/Necesidad real

- La alta dirección desconoce o no está interesada en las iniciativas relacionadas con un modelo de puntuación alternativo.
- Otras prioridades dentro de las áreas de riesgo o análisis.
- No se siente que se tiene un problema grave de tasas de aprobación en el segmento.
- No es una prioridad abordar las iniciativas de inclusión financiera en el segmento.
- Se sienten cómodos con la cantidad de datos (tradicional) que consiguen para tomar una decisión de crédito en el segmento.

13.8. Educación en cultura de empoderamiento, confianza y democratización de datos

La banca pública del desarrollo deberá incorporar en sus programas de educación y acompañamiento temas que involucren la comunicación hacia los micronegocios solicitantes de crédito, respecto a porqué son importantes los datos alterativos y cómo pueden ayudar a contribuir al análisis crediticio, para que se propicie una cultura de entendimiento de que los datos alternativos más que un requisito son un mecanismo para empoderar a estos micronegocios a que puedan tener acceso a mejores ofertas de crédito con diferentes actores. De manera que se perciban los beneficios de esta estrategia de inclusión y democratización del crédito, lo cual permitirá lograr la confianza para que estos micronegocios den permisos para el uso de sus datos.

13.9. Asesoría técnica de apoyo a la banca pública de desarrollo

Como se mencionó anteriormente, la banca de desarrollo es la dueña de la estrategia de inclusión financiera de la economía popular, por lo cual debe actuar como promotora y garante del desarrollo de iniciativas de inclusión. Por este motivo, el Grupo Bicentenario como banca pública de desarrollo, deberá contar con asesoría técnica para asimilar los hallazgos de esta investigación, implementar las recomendaciones y poner en marcha las capacidades que la lleven a integrar la estrategia de diseño e implementación de modelos y procesos de crédito usando datos alternativos para cumplir con la estrategia a la luz de las exigencias del PND.

Debido a la novedad que representa tanto el uso de datos alternativos como el mismo ejercicio de incursionar en este segmento, se sugiere priorizar la asesoría técnica a Bancóldex y a la Banca de las Oportunidades, como principales intervinientes dentro del Grupo Bicentenario para poder llevar a cabo las recomendaciones, en especial las referentes a la implementación de los esquemas de trabajo, que involucran construcción de un programa con múltiples interacciones con aliados y proveedores.

14. Recomendaciones de política pública

14.1. Segmentación de las EPC

Definir y caracterizar detalladamente a una EPC. Esto con el fin de identificarlas como un segmento único, localizado en la base de la pirámide, y con diferencias estructurales y de taxonomía en comparación con las mipymes. Hacer públicas estas definiciones y caracterización servirá para mejorar la asociación de programas sociales y reglamentación de productos financieros ofrecidos a este segmento tan específico.

Implementar identificadores únicos para personas naturales, EPC y mipymes. Cuando las jurisdicciones no cuenten con sistemas eficientes, se pueden establecer alternativas como los números de seguridad social y la identificación fiscal; sin embargo, con el fin de fortalecer el uso e intercambio de datos a escala transfronteriza, en el largo plazo se debe evaluar la viabilidad de implementar un marco consistente como, por ejemplo, el identificador de entidad legal (LEI) del G20 para las empresas o la identificación para el desarrollo (ID4D) del Banco Mundial dirigido a personas (Asobancaria, 2022).

Estas definiciones y caracterizaciones deberán estar alineadas con la segmentación de ruralidad y urbanismo definido por el DNP.

14.2. Definición de producto para pilotear el uso de modelos de crédito alternativo

Al hacer análisis comparativo entre el desempeño de los modelos de crédito, una de las primeras actividades consiste en identificar si el modelo fue usado en el mismo producto y segmento, así como si estaba orientado a un objetivo similar. Aspectos como: destinación del crédito, tipo de crédito (revolvente/no revolvente), monto promedio, monto máximo, plazo promedio, plazo máximo, definición de *default* (por ejemplo, sesenta días de mora en una ventana de observación de seis meses) usualmente son caracterizados.

Para promover la consistencia de los modelos de crédito es ideal que existan condiciones de producto y segmento estables de manera que cuando se realicen las revisiones de desempeño del modelo estas sean comparables a lo largo del tiempo en el mismo modelo, así como con otros modelos.

En ese sentido, se recomienda que para asegurar la consistencia se tomen marcos de regulación financiera, así como estrategia de inclusión financiera de la economía popular y que los incentivos, líneas y productos de garantía sean propuestos con los siguientes cortes:

- **Urbano:** hasta 2 millones COP, con garantía oculta y subsidiada cien por ciento con cobertura hasta el 70 %.
- **Rural:** hasta 4 millones COP, con garantía oculta y subsidiada cien por ciento con cobertura hasta el 70 %.

Dado que el enfoque de la presente investigación ha sido sobre crédito urbano, se recomienda en esa misma línea realizar los primeros pilotos con productos para ese mercado iniciando con una baja exposición de 2 millones COP.

Al considerar los antecedentes de autoexclusión que se mencionaron en capítulos anteriores es importante apuntar productos y soluciones que tengan gran impacto para el micronegocio pero que también sean de su interés; por tal motivo, se recomienda que el préstamo sea para destinación de capital de trabajo, como por ejemplo ampliación de inventario, abastecimiento de insumos, resolver flujo de caja o ampliación de plazos de pago a clientes (Eplee, Kempis, y Ogden, 2023).

Una parte importante de la recuperación del crédito tiene que ver con la capacidad de organizar, no solo el monto y el periodo sino además alinear los ciclos de flujo de caja positivo de recepción de ingresos del micronegocio con el flujo negativo que representaría el pago de la cuota del crédito. Ante esto se recomienda que se pueda pilotear también tener plazos y periodos de amortización semanal o catorcenal, adicional al mensual que es el más común.

Esto contribuirá a evitar retraso en pagos con su correspondiente incumplimiento hacia la IF. A los micronegocios también les proporciona beneficios, en cuanto a planear sus finanzas y mejorar su salud financiera.

Cuando se considera el apalancamiento de iniciativas del Gobierno nacional conviene también tener en consideración la definición del producto para los primeros pilotos, en aspectos financieros y no financieros. Dentro de los financieros, podrían ser incentivos como subsidios a tasas de interés o subsidio a los servicios de fianza de los fondos de garantías y que estos incentivos estén soportados por otros incentivos o estímulos no financieros, como por ejemplo promoción a la igualdad de oportunidades o estímulos de género, acompañamiento y educación financiera a los micronegocios entre otros.

Por último, se recomienda que dentro de la definición de producto, la política de crédito establezca criterios de georreferenciación y concretamente que el micronegocio tenga un lugar fijo establecido para realizar su actividad económica, dícese local, taller o almacén; esto con el fin de maximizar contactabilidad en caso de impago; esta condición se podría ir flexibilizando conforme se van teniendo datos de desempeño de la cartera para ser considerados en la mejora del o de los modelos de crédito alternativo.

14.3. Definición de segmento o sector para pilotear el uso de modelos de crédito alternativo

Los micronegocios del mismo sector industrial tienden a tener modelos de operación comercial similares. En consecuencia, se pueden utilizar tipos similares de datos transaccionales para explorar el estado de sus operaciones comerciales y obtener información sobre su solvencia. La idea de desarrollar modelos basados en sectores o en conglomerados para la calificación crediticia facilita la especialización en el análisis. Este enfoque organiza negocios de naturaleza diversa en múltiples grupos, con negocios en cada grupo que tienen características similares. Esto lleva a que el proceso de calificación crediticia se centre en analizar negocios en el mismo grupo y que tienen puntos de datos similares.

Con base en lo anterior, el financiamiento de la cadena de suministro podría presentar una oportunidad para apoyar a las EPC o micronegocios y a sus clientes ya que utilizar el conocimiento de los proveedores puede resolver uno de los principales desafíos de los préstamos comerciales, que es entender el riesgo crediticio en el contexto de información limitada e incompleta (Eplee, Kempis, y Ogden, 2023). Proporcionar liquidez a los proveedores para mejorar su capacidad de otorgamiento de crédito o recopilar información de los proveedores para suscribir préstamos de capital de trabajo a las propias empresas probablemente también se traspasaría esa liquidez al permitir que estas ofrezcan más crédito al que ya ofrecen.

En la tabla 14-1, se recopilan algunos ejemplos de segmentación por tipo de empresa y datos correspondientes.

Tabla 14-1. Ejemplos de datos transaccionales y no transaccionales por tipo de industria

Sector	Fuente de Datos	Ejemplos de datos transaccionales	Ejemplos de datos no transaccionales
Todos los sectores industriales	Histórico de los records de flujo de efectivo transaccional de las mipymes	Para entrenamiento del modelo: <ul style="list-style-type: none"> • Actividades de Ingresos (<i>revenue</i>) • Actividades de egresos (gastos) • Historial de mora 	N/A
	Estados de cuenta bancarios entregados por las mipymes	Para predicción de impago: <ul style="list-style-type: none"> • Actividades de Ingresos (<i>revenue</i>) • Actividades de egresos (gastos) 	N/A
	Servicios públicos y otros proveedores de datos	Para entrenamiento del modelo y predicción de impago (basada en información de egresos) <ul style="list-style-type: none"> • Historial de consumo (servicios públicos, telefonía, etc.) 	Para entrenamiento del modelo y predicción de mora (basada en información de gastos) <ul style="list-style-type: none"> • Selección de productos suscritos (p.ej. servicios de valor agregado y paquetes) • Solicitudes de pagos diferidos • Domicilio • Categoría de negocio
Sector de <i>retail</i> (minorista)	Histórico de transacciones mediante datáfonos (proveedores de datos de terceros)	Para predicción de impago (basada en ingresos): <ul style="list-style-type: none"> • Valor de transacciones • Frecuencia de transacciones • Promedio diario/semanal/ mensual de transacciones • Tipos de pago 	Para entrenamiento del modelo y predicción de impago <ul style="list-style-type: none"> • Porcentaje de clientes recurrentes • Número de transacciones problemáticas (reembolsos, revertidas) • Información de transacciones riesgosas
	Proveedores del negocio (proveedores de datos de terceros)	Para predicción de impago (basada en egresos): <ul style="list-style-type: none"> • Valor de transacciones de órdenes a proveedores • Frecuencia de transacciones de órdenes a proveedores • Promedio diario/semanal/ mensual de transacciones de las órdenes • Historial de pago a proveedores 	Para entrenamiento del modelo y predicción de impago <ul style="list-style-type: none"> • Estadísticas de proveedores frecuentes • Patrones de solicitudes • Número de transacciones problemáticas (reembolsos, reveridas)

Fuente: Hong Kong Monetary Authority.

14.4. Recomendaciones regulatorias

Una disponibilidad estable de datos alternativos es un requisito previo clave para implementar un score de crédito alternativo, por lo que el apoyo continuo del gobierno es fundamental para garantizar la disponibilidad de datos alternativos. Los reguladores colombianos deberán emitir circulares y decretos que respalden el uso de agregadores *fintech* y firmas de analítica de datos para la gestión del riesgo crediticio, estableciendo los principios que rijan la adopción de nuevas técnicas y prácticas de gestión del riesgo para préstamos de mipymes. Una orientación clara del gobierno sobre el uso y la gestión de datos alternativos facilitará de manera efectiva el intercambio de datos alternativos entre bancos y entidades no bancarias.

Emisión de nuevas iniciativas para promover el uso de datos alternativos

La definición y emisión de nuevas iniciativas respaldadas por el gobierno podrían ayudar a promover el uso de datos alternativos, especialmente si consideran las necesidades de seguridad y privacidad de los datos frente a los beneficios de la innovación *fintech*. Por ejemplo, se podrían permitir usos específicos de datos alternativos para la calificación crediticia bajo condiciones bien definidas. Actualmente, el uso de los datos está regulado por una política general que cuida de los usuarios, pero queda corta en cuanto a promoción del uso responsable de la información.

Definir principios y directrices para la gestión segura de datos alternativos

Definir principios y directrices para la gestión segura de datos alternativos podría ayudar a las IF a usar datos alternativos garantizando seguridad en tres áreas principales:

- **Recopilación, procesamiento y almacenamiento** adecuados de datos alternativos de proveedores de datos de terceros.
- **Consentimiento legítimo de las mipymes** por parte de las IF para la autorización del uso de datos.
- **Aplicación de Machine Learning e IA** de acuerdo con las pautas y principios emitidos, incluidas las regulaciones de privacidad de datos.

14.5. La privacidad como una serie de requerimientos normativos y un eje para el diseño de procesos de crédito y productos financieros

Como se mencionó previamente, el segmento de las EPC y micronegocios tienen un gran componente de informalidad que tiene cierta explicación por el hecho de que estos microempresarios, no tienen la suficiente confianza en las instituciones financieras ni en las autoridades. Una posible consecuencia de esta desconfianza, que requeriría más estudio, podría ser la autoexclusión crediticia (Banca de las Oportunidades, 2022) que se vive en el país.

Otros estudios revelan que el 95 % de las organizaciones B2C (*business to consumer*, por sus siglas en inglés) (empresa a consumidor) cree que los consumidores confían en su capacidad de proteger los datos, pero solo el 65 % de estos confía realmente en estas empresas (Twilio, 2022), es decir existe una brecha de confianza entre los consumidores y las organizaciones; y partiendo de esto, la privacidad en el uso de datos tanto tradicionales como alternativos debe ser puesta al centro como eje principal en el diseño de procesos de solicitud de crédito; y no solo para cumplir con los requerimientos normativos que son obligatorios sino para construir la confianza de este segmento.

La privacidad por diseño es un concepto que incorpora la privacidad desde el inicio y durante todo el ciclo de vida de los procesos, sistemas y servicios que involucran datos personales (Grupo Atico34, s.f.). En el caso de la solicitud de crédito con datos alternativos, la privacidad por diseño es aún más crítica por lo novedosa e “invasiva”, que puede ser considerada la fuente de datos alternativos; por lo que se sugiere realizar un análisis de impacto en la intimidad de los solicitantes para cada fuente de datos seleccionada, identificar los riesgos potenciales y adoptar medidas para prevenirlos o mitigarlos.

14.6. Tratamiento de los datos alternativos

Emitir instrucciones claras sobre el tratamiento que debe darse a los datos alternativos, que aunado a los retos al evaluar las fuentes de datos alternativos mencionados en el capítulo 2 “Contexto sobre las fuentes de datos alternativos”, agravan las dificultades para garantizar la disponibilidad y precisión de esta información. En consecuencia, es necesario realizar las siguientes acciones en política pública:

- Emitir disposiciones regulatorias sobre la forma más adecuada de obtener y procesar esta información.
- Emitir disposiciones regulatorias sobre la vigilancia y facultades de los diferentes actores siendo estas los *fintech*, agregadores de datos y demás proveedores de fuentes de datos alternativos sobre las diferentes modalidades de entrega de datos que le hacen a las entidades financieras y que permita delimitar las responsabilidades entre los agregadores, agencias/firmas de analítica de datos y las entidades financieras que otorgarán los créditos.
- Digitalizar los registros públicos.
- Promover el acceso a sistemas y estándares de datos abiertos para las mipymes y en específico para las EPC (Asobancaria, 2022).

14.7. Fortalecimiento del banco de datos

Aumentar la cobertura de información crediticia tradicional y la de datos alternativos. En este sentido, los reguladores deben:

- Promover el intercambio abierto, justo y competitivo de información crediticia independientemente de si es tradicional o alternativa.
- Redefinir los umbrales respecto al monto mínimo y tipo de producto que deben tener los préstamos para que los burós, agregadores y agencias/firmas de analítica de datos puedan recopilar sus datos de acuerdo con las recomendaciones 1 de segmentación y 2 de segregación. Por ejemplo:

- Si el monto es muy alto puede dejar por fuera mipymes o EPC; y si por el contrario el umbral es muy bajo puede ocasionar fricción a los procesos de crédito.
- Si la definición de producto y segmento no es clara, se pueden llegar a confundir políticas de crédito de consumo y productivo en el segmento de las EPC, que tienen una ambivalencia al ser personas naturales con actividad en cierto sentido empresarial.

14.8. Habilitación de mecanismos de garantía de riesgo compartido

Elaborar mecanismos que habiliten trasladar riesgos o compartición de riesgos. De acuerdo con lo que se ha observado en pilotos de uso de datos alternativos (Arráiz, Bruhn, Ruiz, y Stucchi, 2018), una de las razones para que las entidades financieras estén dispuestas a modelos de *scoring* alternativo es la habilidad de trasladar o compartir el riesgo. Para una IF, implementar modelos de *scoring* alternativo significa testear con nuevas poblaciones, implicando un incremento en el valor en riesgo (VaR), debido a pérdidas asociadas al riesgo adicional. También se espera tener un número de clientes deudores que en un proceso tradicional serían rechazados, si bien es necesario recolectar datos y llegar a un número de “clientes malos” para poder calibrar el nuevo modelo y en un futuro tener un mejor desempeño. Se debe considerar que muchas entidades no podrían aceptar esta variable de incertidumbre. Por tal motivo, incluir mecanismos de garantía de riesgo compartido para cubrir una fracción máxima de la exposición crediticia (principal e intereses) con el fin de obtener préstamos garantizados, es un incentivo para que las entidades financieras participen en los pilotos.

14.9. Uso de datos alternativos para desarrollar productos financieros acorde con las necesidades reales de las EPC

Un estudio de la Financial Access Initiative indicó que existe una brecha entre la necesidad real de productos financieros y la oferta actual de productos de las IF, por ejemplo, la principal necesidad de financiamiento de los micronegocios no es para activos fijos que suponen en la mayoría de los casos una garantía, sino para capital de trabajo, alrededor del 40 % de los micronegocios del estudio utilizaron el financiamiento de la cadena de suministro tales como materia prima, insumos e inventario (Eplee, Kempis, y Ogden, 2023). Otros estudios muestran que esta proporción podría llegar al 86,1 % considerando adicionalmente gastos operativos y de funcionamiento (Banca de las Oportunidades, 2020).

El entendimiento sobre la necesidad y del producto adecuado, puede aumentar la utilización de financiamiento formal en empresas (~30 %), que utilizan financiación formal pero solamente para menos de la mitad de su actividad financiera (Eplee, Kempis, y Ogden, 2023), siendo probable que esto se pueda abordar mediante ajustes en el *marketing* y diseño de productos.

El apalancamiento de datos alternativos, como de la banca y las finanzas abiertas, puede mejorar el acceso a los datos de los actores económicos para que las IF comprendan mejor el comportamiento y los intereses de sus clientes para crear productos y servicios a la medida de sus necesidades (Bignell, 2023).

Si el análisis psicométrico se combina con análisis de data transaccional pública, así como de ciertas redes sociales, se podría disponer de una herramienta muy poderosa tanto para predecir el riesgo como para identificar propensión de consumo, de manera que las estrategias de *marketing* pueden aprovechar también el *big data* e IA para segmentar mejor al cliente y realizar productos a la medida o sugerir productos para segmentos específicos como el de EPC (Monsberger y Del Rosal, 2021).

14.10. Uso de marcos que promuevan la igualdad en las oportunidades de crédito

En análisis de crédito tradicional a menudo incurre en decisiones sesgadas, como por ejemplo raza, color, religión, origen nacional, sexo, estado civil, edad, así como algunas otras fallas de mercado (The United States Department of Justice, 2022). Es por tanto, se recomienda que el uso de datos alternativos siga marcos de trabajo como el de la Ley de Igualdad de Oportunidades de Crédito (ECOA) de los Estados Unidos de América, ya que pueden corregir estos sesgos.

14.11. Uso de marcos que promuevan el buen manejo de los datos alternativos con óptica de central de información

En adición, al cumplimiento de normas y reglamentos de organismos que regulan el manejo de información localmente, como lo serían la Ley 1266 de 2008, también conocida como “Ley de Habeas Data”, o la Ley 1581 de 2012, en las cuales se establece el régimen general de protección de datos personales en Colombia.

Se sugiere que se adopten marcos de regulación o de manejos de información similares a los de la Comisión de Regulación de Agua Potable y Saneamiento Básico (CRA), tal como lo sería la Fair Credit Reporting Act (Ley de Reportes Crediticios Justos) (FCRA); esto significa que es que el proveedor de las fuentes de datos alternativos sea este un fintech, un agregador de *open data*, o agencia/firma de analítica de datos, y para la calificación crediticia debería cumplir con los requisitos y obligaciones de organismos como la FCRA (Federal Trade Commission, 2022). Esto puede incluir obtener el consentimiento del solicitante, garantizar la precisión y privacidad de los datos, responder a las disputas y solicitudes de los consumidores y enfrentar posibles sanciones por incumplimiento.

Glosario de términos

Aprendizaje automático: traducción del inglés *Machine Learning*. Grupo de técnicas que, mediante algoritmos, se pueden analizar gran cantidad de datos y obtener información relevante para la toma de decisiones predictivas en diferentes ámbitos, como en el del riesgo de crédito (Grau Álvarez, 2020).

Archivo delgado: con respecto a la evaluación crediticia, un solicitante de “archivo delgado” es una persona con poco o nulo historial crediticio, lo que significa que no tiene suficiente información en su historial para ser valorado mediante la evaluación crediticia tradicional (por ejemplo, los score de buró).

Agencias o firmas de analítica de datos: son compañías de tecnología y analítica avanzada que poseen amplias capacidades de manejo de datos e información y proporcionan, herramientas, soluciones y SaaS (*software* como servicio), generalmente para apoyar a clientes en la toma de decisiones basadas en datos.

Agregadores de datos: un agregador de datos es una organización que recopila datos de una o más fuentes, proporciona un procesamiento de valor agregado y vuelve a empaquetar el resultado en una forma utilizable. Los proveedores de *Open Finance* generalmente son agregadores.

Backtesting: es un procedimiento general y ampliamente utilizado para evaluar el desempeño del piloto para ciertos tipos de modelos financieros. Es una comparación *ex post* sobre los resultados observados con los resultados esperados derivados del uso de un modelo (*Open Risk*), es decir, se aplican datos históricos y se comparan los resultados previstos con los reales.

Causales de negación “blandas”: también llamados *soft rejects* (rechazos suaves) son las causas por las que un solicitante de crédito no es aprobado de una manera inmediata ya que se encuentra en una “zona gris”, donde el resultado de su análisis crediticio no es lo suficientemente contundente como para aprobarlo o negarlo inmediatamente. Esto se debe generalmente a la falta de información de actividad económica o historial crediticio, por lo cual la solicitud termina siendo negada o desistida, pero no por que se haya confirmado una evidente falta de capacidad de pago o un evidente mal comportamiento de pago, simplemente la falta de información produce incertidumbre y un riesgo asociado, por lo que las IF deciden no otorgar el crédito.

Datos crudos: datos extraídos en su forma original y sin procesar, en ocasiones también llamados “datos de origen”, “datos atómicos” o “datos primarios”. Comúnmente se hace una distinción entre datos e información en el sentido de que esta es el producto final del procesamiento de datos.

Aunque los datos sin procesar tienen el potencial de convertirse en “información”, requieren extracción selectiva, organización y, a veces, análisis y formateo para su presentación. Debido al procesamiento, los datos sin procesar a veces terminan en una base de datos, lo que permite que sean accesibles para su posterior procesamiento y análisis de varias maneras diferentes (Wright, s.f.).

Data Journey o recorrido de datos: representa las etapas clave de un proceso de datos (Statistics Canada, 2021). El recorrido no es necesariamente lineal; tiene la intención de representar los diferentes pasos y actividades que podrían emprenderse para producir información significativa desde la recopilación de los datos crudos, su envío, procesamiento y visualización, pasando por los protocolos de seguridad de la información (The Standard CIO, 2020).

Datos personales: información personal de identificación o datos personales. IPI es cualquier dato que potencialmente podría identificar a un individuo específico.

End point: cualquier elemento tecnológico que proporciona un punto de entrada a los activos y aplicaciones de la organización (CyberArk, s.f.); en contexto de los datos alternativos para análisis de crédito puede ser una dirección URL o un método que permite la interacción o transferencia de datos para obtener los resultados de un solicitante de crédito (ICM, 2021), esto se hace generalmente a través de API.

Economía popular y comunitaria (EPC): término establecido para identificar a la población trabajadora que participa en la economía con microemprendimientos, siendo microempresarios en el segmento principalmente informal y que se está reconociendo formalmente en las bases del Plan Nacional de Desarrollo 2022-2026, dentro de las políticas de inclusión productiva con trabajo decente y apoyo al emprendimiento, en el cual se establece la apertura de líneas de crédito específicas por parte de la banca pública que consideren las condiciones propias de las unidades económicas de la EPC.

Embedded Finance: también denominada “finanzas embebidas”, consiste en la integración de servicios financieros que realizan diferentes plataformas (de cualquier sector) a sus canales digitales para simplificar la experiencia del usuario y evitar la redirección de tráfico hacia otras webs.

Empresa de investigación de audiencias: se define como el conjunto de métodos de investigación utilizados para obtener información sobre la audiencia, y que permite estimar el número de personas que han estado expuestos a un evento medial en un momento del tiempo determinado; este puede referirse por ejemplo al consumo de un producto audiovisual, ya sea textual, sonoro, fotográfico o multimedia. A este tipo de empresas pertenece GWI.

Expediente liviano: véase archivo delgado.

Fintech: es una compañía que utiliza tecnología para proporcionar servicios financieros o facilitar el desarrollo de los mismos, generalmente de manera más eficiente, económica e innovadora que los proveedores tradicionales (CB Insights, 2022). Las instituciones Fintech que ofrecen servicios financieros directos, como lo son pagos móviles, préstamos, inversiones, seguros y asesoramiento financiero; sin embargo, también existen otras que ofrecen servicios indirectos, como por ejemplo datos, *software*, plataformas para que otras fintech u otras IF provean los servicios directamente.

FTPS o FTP sobre SSL: es un nombre que se utiliza para permitir varias formas en las que el *software* FTP puede realizar transferencias de archivos seguras. Cada forma implica el uso de una capa SSL/TLS por debajo del protocolo FTP estándar para cifrar los canales de control o datos (Tecnología Informática, s.f.)

Gini / Coeficiente de Gini: también llamado índice de Gini es un término para evaluar el poder predictivo de un modelo predictivo. Para el caso de un modelo de riesgo crediticio es una indicación sobre la capacidad del modelo para discriminar entre un prestatario “bueno” o “malo”.

El Gini se mide en una escala de 0 a 1, siendo más predictivo un Gini más alto. En la industria crediticia es habitual observar Ginis entre 0,2 y 0,7. Un Gini cambiará con el tiempo y no debe verse como “fijo”. Un modelo de crédito debe ajustarse periódicamente con nueva información de pago, lo que mejorará el Gini.

Informed gaming o “hacer trampa de manera informada” (fuente: elaboración propia de acuerdo con experiencia del equipo consultor): en el contexto de análisis crediticio usando datos alternativos, se tratan de estrategias para falsear información o manipular los resultados, por ejemplo, un puntaje de crédito. Esto se puede tener con diferentes fuentes, como por ejemplo en fuentes de datos de redes sociales o *e-mail* al crear perfiles falsos, publicar información engañosa o manipular su reputación en línea, inducir la proliferación de tráfico en sus correos electrónicos, entre otros para crear una “huella digital artificial”, de la misma manera en fuentes de datos psicométricos sería, por ejemplo, al responder una evaluación psicométrica de una manera que presente un perfil de personalidad o de conducta diferente al real.

Insights: en el contexto de la analítica de negocio y de la inteligencia de negocios, los “*insights*” son el resultado del análisis de los datos y de la información que ayudan a propiciar entendimiento de un tema o problemática y además conocimiento accionable para ser traducido en tácticas o estrategias en el negocio (BiSmart, 2022).

Insights de comportamiento: en el contexto de la evaluación crediticia los *insights* de comportamiento evalúan un conjunto de rasgos de carácter demostrando que predicen la probabilidad de que alguien repague o incumpla. Algunos ejemplos *insights* de comportamiento incluyen: ambición, emprendimiento, optimismo o gestión del tiempo, etcétera. Hay varias formas de obtener información sobre el comportamiento, incluyendo la evaluación de los datos de dispositivos o la evaluación psicométrica (fuente: elaboración propia de acuerdo con experiencia del equipo consultor).

Invisibilidad al crédito: puede usarse para referirse a una persona sin historial crediticio formal, haciéndolo invisible para las instituciones financieras que dependen de puntajes crediticios tradicionales para la evaluación. Sin un historial crediticio, los métodos tradicionales no pueden “ver” a estas personas.

KS o Kolmogorov-Smirnov: indicador estadístico de desempeño de un modelo que mide el poder discriminatorio de este al examinar la diferencia máxima entre la distribución de eventos acumulativos. Para el caso de riesgo de crédito serían eventos de pago (buenos) y no pago (malos). El KS tiene una escala de 0 a 100, en la cual 0 indica que no hay capacidad para predecir resultados buenos o malos, y 100 indica una predicción perfecta (Park, 2020).

Lista OFAC: es un registro emitido por la Oficina de Control de Activos Extranjeros (OFAC) de personas vinculadas con delitos precedentes de lavado de activos, terrorismo, proliferación de armas y su financiación, o cualquier otra amenaza que atente contra la seguridad de los Estados Unidos.

LenddoEFL: una empresa de *software*, con funcionamiento en Singapur.

Marketplace: en el contexto de las finanzas, el modelo *marketplace* es aquel en el cual distintas IF ofrecen sus propios productos y servicios, los cuales se integran el de otras compañías para que el consumidor financiero pueda acceder y elegir dentro de las diferentes soluciones.

Modelo genérico: en el contexto de los modelos estadísticos de riesgo crediticio se refiere a un modelo que no se adapta a una cartera o producto específico, sino que fue construido para aplicarse a una amplia gama de escenarios, productos o geografías de riesgo crediticio. Un modelo genérico generalmente usa datos de todo un sector o mercado para estimar la probabilidad de incumplimiento u otras medidas de riesgo (Jorion, 2007).

Onboarding o Data Oboarding: en el contexto de análisis crediticio, *onboarding* es el proceso de recolección de datos de prospectos o clientes potenciales para productos o servicios crediticios. Implica recopilar datos de diversas fuentes tanto tradicionales como alternativas.

Opt-in: es el proceso de obtener autorización para usar la información de contacto y de las fuentes de datos definidas, a partir de una llamada a la acción de “registro” o de “registro” en la solicitud de crédito misma, que se usará en el futuro para calificar dentro de un modelo de *scoring* o puntaje crediticio.

Puntaje de crédito: *score* de crédito o puntaje de crédito es un análisis de la calidad de crédito de un individuo, por lo general representado por un número de tres dígitos.

Procesamiento de lenguaje natural (PLN): es la práctica del entendimiento de cómo las personas organizan sus pensamientos, sentimientos, lenguaje y comportamiento. Un campo que se extiende hasta las ciencias de la computación, inteligencia artificial y lingüística en el estudio de las interacciones entre las computadoras y los seres humanos. El objetivo es dotar a la máquina de la capacidad de interpretar el texto simulando la habilidad humana de entender el lenguaje (Datacentric, 2019)

Scorecard: es una fórmula que utiliza elementos de datos, o variables, para determinar un umbral de tolerancia al riesgo.

Score de crédito: o puntaje de crédito es un análisis de la calidad de crédito de un individuo, por lo general representado por un número de tres dígitos.

Score de bureau: se refiere a los *scores* o puntajes de crédito que emiten los burós o las centrales de información, los cuales consideran datos tradicionales de historial crediticio reciente de alguien de 5 a 7 años.

SDK (*Software Development Kit*): el kit o estuche de desarrollo de *software* es un conjunto de herramientas proporcionado usualmente por el fabricante de una plataforma de *hardware*, un sistema operativo (OS) o un lenguaje de programación (RedHat, 2020).

Trial: es un periodo de prueba con condiciones especiales. Es una táctica de venta utilizada por las empresas de SaaS, en las que reducen el precio de su producto por un tiempo limitado o como parte de una oferta de lanzamiento o de primer uso con una IF donde se les permite probar la solución (UserPilot, 2022).

Variable: una variable estadística es una característica o cualidad que posee un individuo o un objeto y que puede adoptar diferentes valores, por lo general numéricos, que se pueden medir. Para el caso de los datos demográficos tradicionales podrían ser la edad, el peso, la estatura, el sexo y el nivel de ingresos, etcétera (Economipedia, s.f.). En el contexto de los datos alternativos analizados en el presente documento. Las variables estadísticas resultantes de estas fuentes podrían ser asociadas a actividad en dispositivo móvil, perfil psicométrico, transacciones en monedero electrónico, perfil de redes sociales, consumo de servicios públicos que describan y analicen las propiedades de una población o una muestra de datos.

VaR (*Value at Risk*): El *Credit Value-at-Risk*, se define como la pérdida de crédito no esperada que pudiera sufrirse en un horizonte de inversión y a un nivel de confianza específico. A diferencia del VaR de mercado, que se reporta como el valor asociado al cuantil del nivel de confianza, para el caso de riesgo de crédito, el crédito VaR se reporta como la diferencia entre el valor del cuantil asociado al nivel de confianza y la pérdida esperada de la inversión o portafolio (Consejo Monetario Centroamericano, 2018).

Bibliografía

- Angilella, S. M. (2015). *The financing of innovative SMEs: a multicriteria credit rating model*.
- Arráiz, I., Bruhn, M., Ruiz, C., y Stucchi, R. (2018). Are Psychometric Tools a Viable Screening Method for Small and Medium Enterprise Lending. *Private Sector Series*, TN (5).
- Artarti, R., y Shahnaz, S. (2018). *Psychometric Credit Scoring in Indonesia Microfinance Industry: A Case Study in PT Amartha Mikro Fintek*. ICMEM The 3rd International Conference on Management in Emerging Markets.
- Asociación Bancaria y de Entidades Financieras de Colombia (Asobancaria). (2022). Uso de información alternativa para fortalecer los modelos de scoring. *Banca y Economía* (Edición 1346), 12. https://www.asobancaria.com/wp-content/uploads/2022/09/1346_BE.pdf
- Banca de las Oportunidades. (2020). *Demanda de los productos financieros en los micronegocios colombianos 2019*. Unidad de Análisis Económico.
- Banca de las Oportunidades. (2022). *Encuesta de demanda de inclusión financiera*. Banca de las Oportunidades. Unidad de Análisis Económico.
- BBVA. (2022). *La banca y las finanzas abiertas en Perú*. BBVA.
- Begini. (2022). *Begini. Solutions – Psychometrics*. <https://www.begini.co/psychometrics/>
- Belvo. (enero de 2023). *Productos/Categorización*. Transforma datos brutos en información útil con ayuda del *machine learning*. https://belvo.com/es/productos/categorizacion/?utm_medium=email&_hsmi=245988379&_hsenc=p2ANqtz--2bZZCziHwih22n2Mvejy8JLIZG135_5uzO2FrQ0FnUCj4S_4qjVIPR-_QnD97yUevRVpIfMZ05dTodroxqQFiLBHn2Q&utm_content=245988379&utm_source=hs_automation
- Belvo. (2022). *El estado del open banking en Latinoamérica*. Belvo
- Bignell, F. (22 de febrero de 2023). *The Fintech Times*. Colombia Reaffirms Commitment to Open Banking and Open Finance Development. <https://thefintechtimes.com/colombia-reaffirms-commitment-to-open-banking-and-open-finance-development>
- Birch, D., Cummins, W., y Shin, C. (2018). How Open Banking Can Drive Innovation in Financial Services. *Harvard Business Review*, 96, 18-19.
- BiSmart. (2022). BiSmart Blog. ¿Cuál es la diferencia entre información e *insights*? <https://blog.bismart.com/diferencia-entre-datos-informacion-insights>
- Bocconi, S. B. (2015). Social glass: a platform for urban analytics and decision-making through heterogeneous social data. En *WWW2015Companion, WWW2015Companion* (págs. 175-178). Association for Computing Machinery.
- Burrows, L. (26 de enero de 2023). *Experian Insights*. What is Alternative Credit Scoring? <https://www.experian.com/blogs/insights/2023/01/alternative-credit-scoring/>
- Business Insider. (1 de 6 de 2020). The Rise of Banking-As-A-Service.
- CB Insights. (2022). *Research*. What is fintech? <https://www.cbinsights.com/research/what-is-fintech/>
- Center for Financial Inclusion at Accion and Institute of International Finance. (2018). Accelerating Financial Inclusion with New Data. *Mainstreaming Financial Inclusion: Best Practices Series*, Part 4.
- Colombia Fintech. (marzo de 2022). *Novedades*. ¿Cuáles son y cómo están las billeteras digitales en Colombia? <https://www.colombiafintech.co/lineaDeTiempo/articulo/cuales-son-y-como-estan-las-billeteras-digitales-en-colombia>

Confederación de Cooperativas de Colombia (CONFECOOP). (2023). Gobierno y sector cooperativo impulsarán un ecosistema financiero inclusivo para la economía popular.

<https://confecoop.coop/actualidad/actualidad-2023/gobierno-y-sector-cooperativo-impulsaran-un-e-cosistema-financiero-inclusivo-para-la-economia-popular/>

Consejo Monetario Centroamericano. (2018). *Estimación del riesgo de crédito para portafolios*.

Corr, P., y Plagnol, A. (2018). *Behavioral Economics: The Basics*. Ascent Audio.

Crouhy, M., Galai, D., y Mark, R. (2006). *The essentials of risk management* (1 ed.). McGraw-Hill.

CyberArk. (s.f.). *Seguridad para end points*. <https://www.cyberark.com/es/what-is/endpoint-security/>

Departamento Administrativo Nacional de Estadística (DANE). (2022). *Boletín Estadístico de Empresas 2019-2021*. DANE.

Datacentric. (19 de enero de 2019). *Business Blog*. ¿Qué es el procesamiento del lenguaje natural y por qué es la revolución del futuro?

<https://www.datacentric.es/blog/business/procesamiento-lenguaje-natural-revolucion-futuro/>

Departamento Nacional de Planeación (DNP). (2023a). *Plan Nacional de Desarrollo Colombia*. Proyecto de Ley - Articulado PND 2022-2026:

[https://colaboracion.dnp.gov.co/CDT/portalDNP/PND-2023/2023-02-07-Poyecto-del-ley-PND-2022-2026-\(Articulado\).pdf](https://colaboracion.dnp.gov.co/CDT/portalDNP/PND-2023/2023-02-07-Poyecto-del-ley-PND-2022-2026-(Articulado).pdf)

Departamento Nacional de Planeación (DNP). (2023b). *Plan Nacional de Desarrollo de Colombia*. Principales metas del PND 2023-2026:

<https://colaboracion.dnp.gov.co/CDT/portalDNP/PND-2023/2023-02-23-METAS.pdf>

Departamento Nacional de Planeación (DNP). (2023c). *Lineamientos de política pública para mejorar los modelos de calificación crediticia (scoring) para población vulnerable, jóvenes y mipymes utilizando información alternativa*. Dirección de Innovación y Desarrollo empresarial.

Dietrich, L., De Souza, F., y Guerreiro, A. (2020). *Development of credit scores with telco data using ML and agile methodology in Brazil*, 4831. S Global Forum 2020.

Economipedia. (s.f.). *Economipedia*. Variable estadística. Qué es, tipos y ejemplos.

<https://economipedia.com/definiciones/variable-estadistica.html>

EFL. (2017, 7 de marzo). *Cómo EFL ayuda a Te Creemos a aprobar más personas*.

<https://youtu.be/c0b57p9lnm8>

Eplee, R., Kempis, M., y Ogden, T. (2023). *Small Firm Diaries*. | *Colombia: Estudio País*. NYU Wagner. Financial Access Initiative.

Equifax. (2019). *Knowledge Centre*. Open banking and Credit Risk: Friend or Foe?

<https://www.equifax.co.uk/resources/open-banking-credit-risk-friend-or-foe.html>

Esterkin, J. D. (2008, 13 de agosto). *Project Management Institute*. Obtenido de Simulación del manejo de la triple limitación en la ejecución del proyecto.

<https://www.pmi.org/learning/library/es-simulacion-de-gestion-de-proyectos-de-triple-restriccion-7091>

Exabel. (2022, 6 de septiembre). *Blog*. The end-to-end challenges of evaluating alternative data.

<https://www.exabel.com/blog/the-end-to-end-challenges-of-evaluating-alternative-data>

Experian. (2018). *Experian News*. The Impact of open banking on Credit Risk Assessment.

<https://www.experianplc.com/media/news/2018/the-impact-of-open-banking-on-credit-risk-assessment/>

Federal Trade Commission. (2022, agosto). *Fair Credit Reporting Act*. Legal Library. <https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act>

Fair, Isaac and Company (FICO). (1 de febrero de 2022). *FICO Blog*. Using Alternative Data in Credit Risk Modelling. <https://www.fico.com/blogs/using-alternative-data-credit-risk-modelling>

Goodman, L. (17 de June de 2014). The rise of big data credit scoring. *BBC News*.

Google Play. (s.f.). *Google Developers*. Recuperado el 2022, de Desarrolladores de Android/ Documentos/Guías. <https://developer.android.com/guide/topics/data>

Grau Álvarez, J. (2020). *Machine Learning y riesgo de crédito*. Universidad Pontificia Comillas, Facultad de Ciencias Económicas y Empresariales.

Grupo Atico34. (s.f.). *Protección de datos*. Privacidad desde el diseño y por defecto (Privacy by Design). <https://protecciondatos-lopdc.com/empresas/privacy-by-design/>

Grupo Bicentenario. (22 de febrero de 2023). *Noticias*. Lucha contra el gota a gota e impulso a la inclusión la hará G. Bicentenario. <https://www.grupobicentenario.gov.co/prensa/>

Harvard Business School Alumni. (2016). *It's all about who you know: Lenddo makes credit decisions based on your social network*. Harvard Business School, Digital Initiative. <https://d3.harvard.edu/platform-rctom/submission/its-all-about-who-you-know-lenddo-makes-credit-decisions-based-on-your-social-network/>

Hong Kong Applied Science and Technology Research Institute. (2020). *Alternative Credit Scoring of Micro, Small and Medium Sized Enterprises*. Hong Kong Monetary Authority. ASTRI Publications. <https://www.astri.org/wp-content/uploads/2016/11/White-paper-on-credit-scoring.pdf>

ICM. (junio de 2021). *¿Qué son los end points y para qué sirven?* <https://www.icm.es/2021/06/15/que-son-endpoints/>

Jorion, P. (2007). *Financial risk manager handbook*. John Wiley & Sons.

La República. (23 de abril de 2022). *¿Cómo el open banking puede coexistir con la ley de protección de datos en Colombia?* *Technocio*, pág. 1.

LenddoEFL. (27 de marzo de 2018). *Blog*. Score Confidence: Boosting Predictive Power. <https://lenddoefl.com/news/2018/5/27/blog-score-confidence-boosting-predictive-power>

LexisNexis. (2018). *Case Study*. Finance company leverages New Scoring solution to approve more Small Business loan. <https://risk.lexisnexis.com/insights-resources/case-study/finance-company-leverages-new-scoring-solution>

Martinez, E., y Torres, J. (febrero de 2020). *Espacio controlado de prueba para actividades de innovación*. MinHacienda, Unidad de Regulación Financiera. https://www.urf.gov.co/webcenter/ShowProperty?nodeId=%2FConexionContent%2FWCC_CLUSTER-124021%2F%2FidcPrimaryFile&revision=latestreleased

McKinsey y Company. (2021). *Financial services unchained: The ongoing rise of open financial data*. McKinsey.

Ministerio de Hacienda y Crédito Público. (25 de julio de 2022). *Decreto 1297 de 2022*. República de Colombia.

Monsberger, R. J., y Del Rosal, V. (octubre de 2021). *Sparkassenstiftung Latinoamerica*. El futuro de la inclusión financiera: tecnologías y estrategias Fintech para microfinanzas inclusivas y sustentables. https://sparkassenstiftung-latinoamerica.org/fileadmin/user_upload/pdf/Estudios/El_Futuro_de_la_Inclusi%C3%B3n_Financiera_Tecnolog%C3%ADas_y_Estrategias_Fintech_para_Microfinanzas_Inclusivas_y_Sustentables_versi%C3%B3n_final.pdf

Ministerio de Hacienda y Crédito Público. (25 de julio de 2022). *Decreto 1297 de 2022*. República de Colombia.

Monsberger, R. J., y Del Rosal, V. (octubre de 2021). *Sparkassenstiftung Latinoamerica*. El futuro de la inclusión financiera: tecnologías y estrategias Fintech para microfinanzas inclusivas y sustentables. https://sparkassenstiftung-latinoamerica.org/fileadmin/user_upload/pdf/Estudios/El_Futuro_de_la_Inclusi%C3%B3n_Financiera_Tecnolog%C3%ADas_y_Estrategias_Fintech_para_Microfinanzas_Inclusivas_y_Sustentables_versi%C3%B3n_final.pdf

Mora, K. V. (23 de enero de 2023). La penetración de billeteras digitales en Colombia alcanzó a 76 % de la población. *La República*.

Nettle, D. (2020). *Personality: What Makes You the Way You Are*. HighBridge Audio.

Nuñez Hurtado, J. D. (19 de septiembre de 2016). *Accion*. Enabling financial inclusion for the missing middle. <https://www.accion.org/missing-middle>

Open Risk. (s.f.). *Open Risk Manual*. Backtesting: <https://www.openriskmanual.org/wiki/Backtesting>

Park, S. (18 de noviembre de 2020). *Experian*. Understand and Use a Business Credit Risk Score – 5-minute FUNdamentals. <https://www.experian.com/blogs/business-information/2020/11/18/understand-and-use-a-business-credit-risk-score-5-minute-fundamentals/#:~:text=The%20KS%20is%20just%20the,predict%20credit%20risk%20at%20all>.

Plaitakis, A., y Staschen, S. (2020). *Open banking: How to Design for Financial Inclusion*. CGAP/World Bank.

Popper, B. (30 de septiembre de 2014). LendUp Says It Has a Better Alternative to Payday Loans. *The New York Times*.

Putra, S. G. (2018). *SME Credit Scoring Using Social Media Data*. Delft, Delft University of Technology.

Putra, S., Bikash, J., Redi, J., y Bozzon, A. (2020). A Credit Scoring Model for SMEs Based on Social Media Data. *Web Engineering 2020*, 113-129.

Qualco. (2022). *Open banking: Unlocking the collections value chain*. Qualco.

RedHat. (20 de junio de 2020). *RedHat*. ¿Qué es un SDK? <https://www.redhat.com/es/topics/cloud-native-apps/what-is-SDK#:~:text=elegir%20Red%20Hat%3F-%C2%BFQu%C3%A9%20es%20un%20kit%20de%20desarrollo%20de%20software%20o%20SDK,o%20un%20lenguaje%20de%20programaci%C3%B3n>.

Rincón, E. (2023). *Mitos y leyendas del Sandbox ¿Mentira o posibilidad?* https://www.linkedin.com/posts/erick-rincon-cardenas-06134248_mitos-y-leyendas-del-sandbox-mentira-o-activity-7031572031632470016-4Esu?utm_source=share&utm_medium=member_desktop

Roa, L., Correa-Bahnsen, A., y Rodríguez-Rey, A. (2021). *Supporting Financial Inclusion with Graph Machine Learning and Super-App Alternative Data*.

Rojas, A. (septiembre de 2014). *Prezi*. <https://prezi.com/mkefloqrmsey/seguimiento-efl/>

SBS. (2022). Banca abierta: un reto para la innovación en el sistema financiero. *Boletín Semanal: SBS Informa*.

Secretaría Distrital de Desarrollo Económico. (8 de marzo de 2023). *Bogotá productiva local*. Alcaldía de Bogotá lanza la plataforma Crédito Peso a Peso para apoyar la inclusión financiera de la economía popular de Bogotá.

Bogotá.<https://desarrolloeconomico.gov.co/alcaldia-de-bogota-lanza-la-plataforma-credito-peso-a-peso-para-apoyar-la-inclusion-financiera-de-la-economia-popular-de-bogota/>

Statistics Canada. (2021). *Data literacy training*. Data journey.

<https://www.statcan.gc.ca/en/wtc/data-literacy/journey>

Swacha-Lech, M., y Solarz, M. (2019). *Mental accounting in the context of savings and credit decisions. Evidence from Poland*. Argumenta Oeconomica.

https://www.dbc.wroc.pl/Content/73324/Swacha-Lech_Solarz_Mental_accounting_in_the_context.pdf

Tecnología Informática. (s.f.). *Transferencias seguras*. ¿Qué es FTP? Significado. FTPS y SFTP:

https://www.tecnologia-informatica.com/que-es-ftp/#FTPS:_Transferencias_seguras

Thaler, R. H., y Sunstein, C. R. (2021). *Nudge: The Final Edition*. Penguin Lcc Us.

The Standard CIO. (2020). *Información 360° estrategia*. *El data journey: ¿Qué es y hacia dónde nos llevan los datos?*

<https://thestandardcio.com/2020/03/02/el-data-journey-que-es-y-hacia-donde-nos-llevan-los-datos/>

The United States Department of Justice. (18 de noviembre de 2022). *Civil Rights Division*. The Equal Credit Opportunity Act (ECOA). <https://www.justice.gov/crt/equal-credit-opportunity-act-3>

The World Bank Group. (2022). *Key Principles for Effective Regulation and Supervision of Credit Reporting Service Providers*. International Committee on Credit Reporting. World Bank Publications.

Twilio. (2022). *Informe del estado de la interacción con los clientes 2022*.

<https://www.twilio.com/es-mx/state-of-customer-engagement>

United Nations Development Programme Pakistan. (Junio de 2022). Democratizing lending in Pakistan. *Development advocate Pakistan*, 9(2), 50-54.

Universidad Externado de Colombia. (s.f.). *Derecho*. Retos del *Sandbox* Regulatorio en Colombia.

<https://www.uexternado.edu.co/derecho/retos-del-sandbox-regulatorio-en-colombia/>

Urjanet. (2019). *Building a Risk Model with utility payment data*.

https://gob.urjanet.com/rs/426-CYT-566/images/Building%20a%20Risk%20Model%20with%20Utility%20Payment%20Data%20-%20Urjanet.pdf?mkt_tok=NDI2LUdZVC01NjYAAAGIsJketC0yniSPhZ14UPqFgYjw1SVzKvrgP-ahouN7KKYEofp3lc1QIXjuEqKK0puLAKx9f5K6OOrbjF-hoSqKn6F0qi99AzyP1NaJv

UserPilot. (18 de febrero de 2022). *User Pilot*. Blog/Growth/SaaS:

<https://userpilot.com/blog/saas-trial-pricing-examples-best-practices/>

Viani B. Djeundje, J. C. (2020). Enhancing Credit Scoring with Alternative Data. *Expert Systems with Applications*.

Wright, G. (s.f.). *TechTarget*. Data Governance - Database management - Raw Data.

<https://www.techtarget.com/searchdatamanagement/definition/raw-data>



BANCO INTERAMERICANO DE
DESARROLLO - **BID**

DEPARTAMENTO NACIONAL DE
PLANEACIÓN - **DNP**