



Escuela de Posgrados

Caso E: Analítica de Datos y Estrategias de Transformación Digital

**Trabajo de titulación presentado como requisito para optar al título de Magíster en
Gerencia de Datos y Negocios**

Autor:

Carolina Neira Tinoco

Director:

Marco Reyes Clavijo

Cuenca, Ecuador 2026

AGRADECIMIENTOS

Quisiera expresar mi más profundo agradecimiento al Ing. Paúl Ochoa, Vicepresidente de Transformación Digital, quien me abrió las puertas para aplicar todos los conocimientos adquiridos en este programa de estudios y aportar de manera significativa al crecimiento de la organización.

Un agradecimiento especial a mi esposo por su constante apoyo a lo largo de todo este trayecto. A mis padres por motivarme a impulsar mi carrera profesional. Asimismo, un agradecimiento a mis amigos y compañeros que aportaron significativamente con su retroalimentación continua durante este aprendizaje.

RESUMEN

Este proyecto de tesis abordó la necesidad de contar con estrategias eficaces de analítica de datos enfocadas en potenciar un producto denominado *Billetera Digital*, el cual se desarrolló en un mercado altamente volátil donde varios medios de pago digitales disputaron su posicionamiento. A partir de la identificación de los desafíos en la recolección, procesamiento, visualización y gobernanza de datos, se propuso un enfoque integral que combinó herramientas de *Big Data* y principios de transformación digital para optimizar la toma de decisiones. La investigación, de enfoque cuantitativo y alcance descriptivo, utilizó estudios de caso, análisis estadístico y visualización de datos para evaluar patrones, comportamientos y tendencias relevantes. Los resultados contribuyeron a la implementación de modelos de visualización claros, sostenibles y estratégicos, así como al fortalecimiento de capacidades analíticas en la empresa. Con ello, se aportó tanto al conocimiento académico como a la aplicación profesional, facilitando decisiones basadas en datos confiables y oportunamente presentados.

PALABRAS CLAVE

Analítica de datos, *billetera digital*, transformación digital, *Big Data*, segmentación de clientes, gobernanza de datos.

ABSTRACT

This thesis project addressed the need for effective data analytics strategies focused on enhancing a product called *Digital Wallet*, which operated in a highly volatile market where various digital payment methods competed for positioning. From the identification of challenges in data collection, processing, visualization, and governance, a comprehensive approach was proposed that combined *Big Data* tools and digital transformation principles to optimize decision-making. The research, with a quantitative approach and descriptive scope, used case studies, statistical analysis, and data visualization to evaluate relevant patterns, behaviors, and trends. The results contributed to the implementation of clear, sustainable, and strategic visualization models, as well as to the strengthening of analytical capabilities in the company. Thus, it contributed to both academic knowledge and professional application, facilitating decisions based on reliable and timely presented data.

KEYWORDS

Data analytics, *digital wallet*, digital transformation, *Big Data*, customer segmentation, data governance.

A handwritten signature in blue ink, appearing to be 'Neira', is located in the bottom right corner of the page. The signature is stylized and written over a faint, diagonal line.

ÍNDICE DE CONTENIDO

Contenido

INTRODUCCIÓN	1
Estado del arte y Marco Teórico	1
Objetivo general	2
Objetivos específicos.....	2
BIG DATA SOLUTIONS.....	4
Contexto del mercado digital en Ecuador	4
Mercado <i>e-commerce</i> y costos asociados	6
PROCESAMIENTO Y VISUALIZACIÓN DE DATOS	8
Base de datos - Personas	8
Atributos contenidos en la base personas	8
Proceso de limpieza y preparación de la base de personas	9
Base de datos - Comercios	10
COMPETITIVIDAD Y TRANSFORMACIÓN DIGITAL.....	13
Aplicación de la Matriz PEYEA	13
Matriz FODA.....	13
Calificación y ponderación de variables	13
Ubicación en el cuadrante PEYEA	15
Definición de estrategias	16
GESTIÓN Y DISEÑO DE PROYECTOS BIG DATA	18
Objetivo general del proyecto.....	18
Objetivos del negocio	18
Objetivos de la analítica avanzada.....	18
Criterios de éxito.....	18
Modelado	18
Modelo no supervisado: <i>Clustering</i>	18
Modelo supervisado: Predicción de activación.....	22
Resultados esperados e impacto estratégico.....	24

ECONOMÍA Y FINANZAS PARA LA GESTIÓN DE PROYECTOS	25
Contexto de rentabilidad.....	25
Proyección financiera	25
Indicadores financieros VAN y TIR	25
Estructura de tarifas <i>MDR</i>	26
CONCLUSIÓN	27
Recomendaciones estratégicas	27
REFERENCIAS BIBLIOGRÁFICAS	28
ANEXOS	30
Anexo 1. Capturas del tratamiento de datos	30
Anexo 2. Índice de Davies-Bouldin - Selección de <i>clústers</i>	31
Anexo 3. Tabla de centroides - Personas	32
Anexo 4. Modelos predictivos	33
Anexo 8. Proceso global en RapidMiner.....	37

ÍNDICE DE FIGURAS

Figuras

Figura 1 Métricas Clave de Conectividad Digital en Ecuador	4
Figura 2 Evolución del E-commerce en Ecuador	5
Figura 3 Retención de Usuarios Activos	10
Figura 4 Evolución por Monto de Transacción	10
Figura 5 Distribución de Comercios por Estado	12
Figura 6 Tipo de Comercio por monto transaccionado	12
Figura 7 Matriz PEYEA.....	16

ÍNDICE DE TABLAS

Tablas

Tabla 1 Métricas fundamentales de conectividad digital en Ecuador	6
Tabla 2 Comparativa de costos por plataforma de venta	7
Tabla 3 Ecosistema de billeteras digitales en Ecuador	7
Tabla 4 Estadística Descriptiva Base Personas	8
Tabla 5 Estadística Descriptiva Comercios	11
Tabla 6 Matriz FODA	13
Tabla 7 Calificación y ponderación variables PEYEA	14
Tabla 8 Cálculo de Coordenadas Matriz PEYEA	15
Tabla 9 Características Segmento Alto Valor	22
Tabla 10 Proyección Financiera Billetera Digital	25
Tabla 11 Tarifa por Segmento de Comercio	26

INTRODUCCIÓN

Problemática

En la actualidad, las organizaciones enfrentan un entorno altamente volátil, lo que exige desarrollar destrezas y capacidades para una toma de decisiones eficaz. Uno de los principales retos radica en la obtención de información precisa, confiable y oportuna, así como en su correcta manipulación y análisis.

El proceso de recolección, procesamiento y adecuación de datos representa una de las fases más complejas de este ciclo, ya que de él depende la calidad de la interpretación posterior. El manejo de altos volúmenes de datos requiere un tratamiento sistemático y estructurado desde la fuente, permitiendo al usuario final identificar patrones, comportamientos y tendencias tanto en el entorno interno como externo.

Este contexto evidencia la necesidad de establecer una adecuada gobernanza de datos, que incluya políticas claras, roles definidos, procesos, estándares y controles que garanticen que los datos sean consistentes, seguros y alineados a la estrategia organizacional. Sin estos elementos, el proceso de análisis pierde valor y se incrementa el riesgo de decisiones mal fundamentada.

Además, incluso cuando los datos han sido correctamente procesados, una visualización deficiente o confusa puede comprometer la interpretación y generar errores significativos en la apreciación. Una representación gráfica poco clara puede provocar percepciones erróneas del escenario analizado, afectando negativamente su desempeño. Muchas organizaciones fallan precisamente en este punto crítico del proceso donde una mala impresión puede desencadenar consecuencias graves en distintos niveles (Benavides Gil, Martínez Rodríguez, & Segura Ortiz, 2024).

Estado del arte y Marco Teórico

Hoy en día, las organizaciones se enfrentan a un entorno empresarial dinámico, caracterizado por cambios constantes, ciclos de vida de productos más cortos y una alta presión por responder rápidamente a las necesidades del mercado. Este escenario ha impulsado a muchas empresas a adoptar herramientas como la analítica de datos, el Big Data y la inteligencia artificial como parte de su estrategia de gestión. Según Micus et al. (2023), estas tecnologías no solo fortalecen la competitividad, sino que permiten tomar decisiones más informadas y alineadas con las expectativas del cliente. A través de la analítica de datos, las empresas pueden interpretar mejor la información que deja de ser simples registros para convertirse en activos estratégicos que guían el rumbo del negocio.

En este contexto, la inteligencia artificial aplicada a la analítica de datos ha ganado un papel prioritario en la gestión empresarial. Tecnologías como el machine learning y el deep learning han demostrado ser herramientas altamente efectivas para mejorar aspectos clave del negocio como el diseño de productos, la satisfacción del cliente y la entrega de valor (Zhuanghao et al., 2025).

Lo interesante de estos enfoques es su capacidad para aprender y adaptarse a partir de grandes volúmenes de datos, lo que permite anticipar comportamientos, segmentar mercados y diseñar estrategias más acertadas. En el caso del deep learning, por ejemplo, su uso de redes neuronales permite comprender relaciones complejas entre variables que antes eran difíciles de detectar. Esto se traduce en predicciones más precisas y, por lo tanto, en decisiones más sólidas. Kraus et al. (2020) señalan que este tipo de modelamiento resulta especialmente valioso para interpretar conductas del cliente y generar una ventaja competitiva sostenida.

Además del enfoque analítico, el componente estratégico de la sostenibilidad ha cobrado una importancia creciente en la gestión de productos y servicios. Las empresas ya no solo buscan satisfacer al cliente, sino también cumplir con estándares ambientales y sociales. La analítica de datos ha resultado ser una herramienta clave para entender las preferencias del consumidor en relación con productos sostenibles, así como para proyectar el ciclo de vida de estos bienes (Woon et al., 2025). Esta información permite tomar decisiones más responsables y alineadas con los principios de la economía circular.

Objetivo general

Desarrollar un modelo estratégico de analítica de datos que integre herramientas de procesamiento, visualización y gobernanza de la información para fortalecer la toma de decisiones en procesos de transformación digital organizacional referente al posicionamiento de una nueva billetera digital en el mercado.

Objetivos específicos

- Identificar herramientas adecuadas para la manipulación y manejo de grandes volúmenes de datos.
- Representar de manera gráfica los resultados obtenidos mediante técnicas de visualización efectivas.
- Segmentar personas y comercios que hayan transaccionado al menos una vez en la billetera digital, identificando patrones de comportamiento y uso.
- Desarrollar un modelo de predicción de usuarios activos que permita identificar potenciales usuarios con alta probabilidad de activación, tanto dentro del portafolio actual de la billetera digital como en la base de clientes de la entidad financiera que aún no utilizan la billetera.
- Evaluar la viabilidad financiera del proyecto mediante proyecciones de flujo de caja, valor actual neto (VAN) y tasa interna de retorno (TIR).

Una entidad financiera ecuatoriana reconocida lanzó al mercado en octubre 2025 su primera billetera digital. En conjunto con alianzas estratégicas con el gobierno local, se buscó generar alto impacto en su lanzamiento con una propuesta de valor que permite a sus usuarios pagar todas sus diligencias municipales a través de esta billetera, además de poder pagar el transporte público y hacer pagos P2P. Sin embargo, existen competidores fuertes en el mercado que llevan más de un año posicionándose. El primer objetivo clave es generar adquisición de usuarios y comercios en el austro debido al posicionamiento predominante de esta entidad financiera, la cual se encuentra en un ranking de posicionamiento 1-5 entre los demás bancos, con este antecedente la ventaja competitiva es alta para posicionarse en el mercado. Con 4 meses en el mercado se tiene una activación del 23% para usuarios y del 18% en comercios.

Si bien la billetera digital ha logrado una adopción inicial relevante en términos de afiliaciones, este crecimiento no se ha traducido de manera consistente en activación ni uso transaccional recurrente. La principal problemática identificada es que los esfuerzos de adquisición actuales no se encuentran plenamente direccionados hacia perfiles con alto potencial de uso, lo que limita la eficiencia de la fuerza comercial y reduce la conversión posterior a la afiliación.

En este contexto, resulta crítico identificar qué tipo de personas y comercios utilizan efectivamente la billetera digital, cuáles son sus características diferenciales y cómo estos patrones pueden ser replicados tanto en clientes actuales de la entidad financiera como en nuevos prospectos.

BIG DATA SOLUTIONS

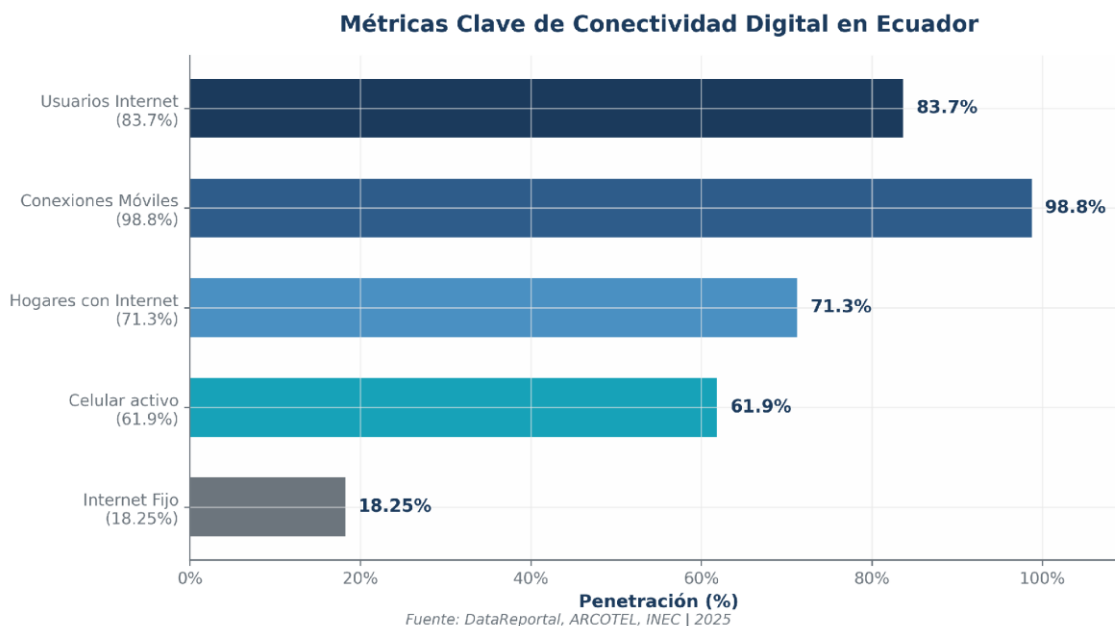
Contexto del mercado digital en Ecuador

Ecuador presenta una conectividad digital consolidada con 15.2 millones de usuarios de Internet (83.7% de la población) (DataReportal, 2025). El comercio electrónico cerró 2024 con ventas cercanas a USD 5,500 millones, con proyección de aproximadamente USD 6,500 millones para 2025 (CECE, 2024; PCMI, 2024). Sin embargo, persisten brechas estructurales: los costos logísticos suman aproximadamente USD 4 por pedido (alrededor del 17% del valor) (Estudios sectoriales *e-commerce* Ecuador, 2024-2025) y los *marketplaces* locales cobran comisiones del 10-20% (Mercado Libre retiene aproximadamente 13-16% según categoría) (Mercado Libre, 2025). Plataformas emergentes como *Marketplace* Ecuador han iniciado modelos de 0% comisión, generando presión competitiva sobre las tarifas tradicionales. (DataReportal, 2025; ARCOTEL, 2025; INEC, 2025)

En la Ilustración 1 se observa que Ecuador exhibe un ecosistema digital consolidado, donde la penetración de Internet móvil (98.8%) supera ampliamente la fija (18.25%), lo que orienta la estrategia hacia soluciones *mobile-first*. El 71.3% de hogares conectados y 61.9% de población con celular activo confirman una base sólida para el crecimiento del *e-commerce* y pagos digitales (DataReportal, 2025; ARCOTEL, 2025; INEC, 2025)

Figura 1

Métricas Clave de Conectividad Digital en Ecuador

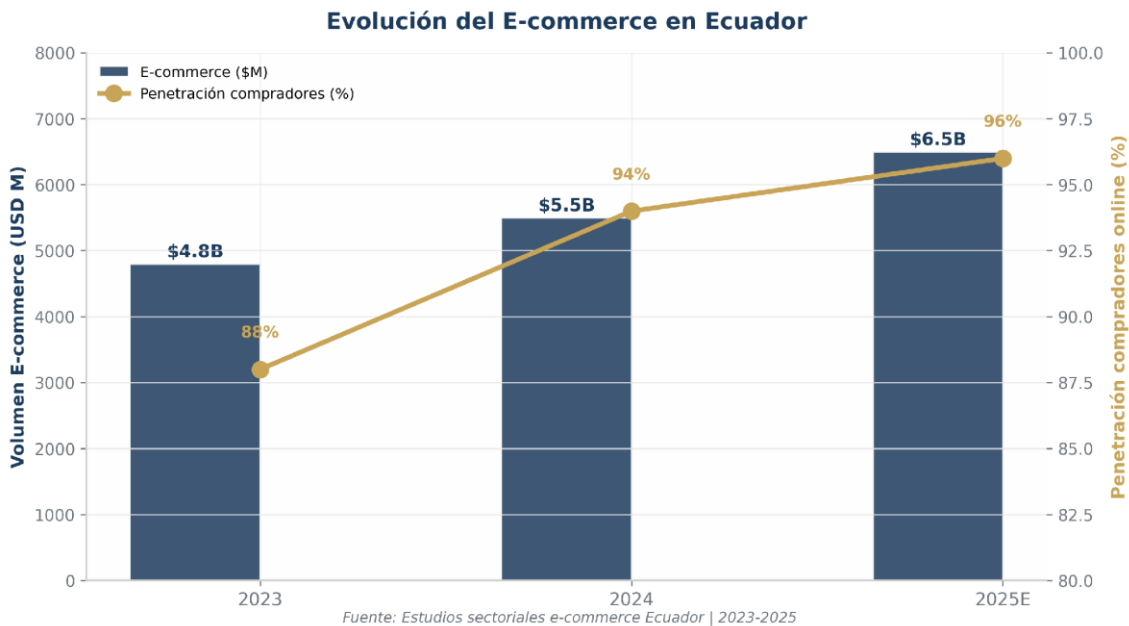


Las billeteras digitales en Ecuador operan en un ecosistema fragmentado y sin interoperabilidad (Banco Central del Ecuador, 2025). Entre las plataformas activas destacan: la billetera digital de la entidad financiera en estudio, PayPhone (Movilnet/Cooperativas), Deuna (Pichincha), PeiGo (Guayaquil), BIMO (Banred/Pichincha), WIP (Produbanco) y Panapay (Coonecta). Ninguna domina el mercado y la Superintendencia de Bancos no reporta cifras agregadas de usuarios o volumen.

La estrategia para la billetera digital debe capitalizar la masiva conectividad (83.7% penetración Internet) y el gran mercado *e-commerce* (aproximadamente USD 5.5B) para posicionarse como opción preferida de pago digital en Ecuador, mediante cuatro pilares: *Adquisición, Retención, Pricing y Alianzas Logísticas*.

Como se ilustra en la **Figura 2**, la evolución del *e-commerce* en Ecuador muestra una trayectoria de crecimiento sostenido que respalda la oportunidad de mercado para la billetera digital (CECE, 2024).

Figura 2
Evolución del E-commerce en Ecuador



Ecuador exhibe un ecosistema digital consolidado. La penetración de Internet móvil (98.8%) supera ampliamente la fija (18.25%), lo que orienta la estrategia hacia soluciones *mobile-first*. El 71.3% de hogares conectados y 61.9% de población con celular activo confirman una base sólida para el crecimiento del *e-commerce* y pagos digitales. La Tabla N°1 presenta las métricas fundamentales

de conectividad digital en Ecuador, las cuales constituyen la base para el desarrollo de estrategias de pagos digitales:

Tabla 1

Métricas fundamentales de conectividad digital en Ecuador

Métrica	Valor	Fecha
Usuarios de Internet	15.2 millones (83.7%)	Ene 2025
Conexiones Internet fijo	3,303,151 cuentas	Sep 2025
Conexiones móviles	18.0 millones (98.8%)	Ene 2025
Celulares con servicio active	61.9% población	Jul 2025
Hogares con Internet	71.3%	Jul 2025

Nota. Elaboración propia con datos de DataReportal, ARCOTEL e INEC (2025)

Mercado e-commerce y costos asociados

Indicadores clave del e-commerce ecuatoriano:

- Tamaño del mercado 2024: aproximadamente USD 5,500 millones (todos los medios de pago)
- Proyección 2025: aproximadamente USD 6,500 millones (crecimiento aproximadamente 18%)
- Penetración de compradores online: 94% de internautas encuestados (2024)
- Ticket promedio estimado: USD 50-54 por pedido en portales locales
- Comisión Mercado Libre: 13-16% según categoría (Clásica/Premium)
- Comisión Marketplace Ecuador: 0% (modelo disruptivo)
- Costo logístico: aproximadamente USD 4-5 por pedido (aproximadamente 10-20% del monto vendido)

El *e-commerce* ecuatoriano muestra un mercado maduro entre usuarios conectados (94% de internautas compran *online*) (CECE, 2024), pero el *e-commerce* como porcentaje del PIB sigue siendo bajo (<3%), indicando espacio significativo para mayor adopción. La presión competitiva sobre comisiones es evidente: mientras Mercado Libre mantiene tarifas del 13-16%, Marketplace Ecuador opera con 0% comisión, democratizando el acceso para emprendedores (Estudios sectoriales e-commerce Ecuador, 2024-2025). El costo logístico (aproximadamente USD 4-5/pedido) sigue siendo la principal barrera de fricción para el crecimiento sostenido.

La fragmentación del ecosistema de ventas *online* en Ecuador presenta una oportunidad clara para la billetera digital. Las plataformas con 0% comisión (Marketplace Ecuador, TikTok, Facebook) están ganando tracción entre emprendedores, pero carecen de soluciones de pago

integradas robustas. Mercado Libre, aunque dominante, enfrenta presión por sus tarifas del 13-17% + envío. La billetera digital puede posicionarse como el medio de pago preferido para estas plataformas emergentes, ofreciendo comisiones competitivas y *experiencia de pago sin fricción*. Como se presenta en la **Tabla N° 2**, la comparativa de costos por plataforma de venta evidencia la fragmentación del mercado y las oportunidades de posicionamiento para la billetera digital.

Tabla 2

Comparativa de costos por plataforma de venta

Plataforma	Comisión	Costo Logístico	Modelo
Mercado Libre (Clásica)	13%	~15% del pedido	Comisión + envío
Mercado Libre (Premium)	17%	~15% del pedido	Comisión + envío + visibilidad
Marketplace Ecuador	0%	Variable (vendedor)	0% comisión + planes
TikTok Live Shopping	0%	~5% del pedido	Livestream + link de pago
Facebook Marketplace	0%	Variable (vendedor)	Orgánico + contacto directo

Nota. Elaboración propia con datos de DataReportal (2025)

El ecosistema de billeteras digitales en Ecuador es altamente fragmentado. Cada plataforma opera en "silos" sin interoperabilidad entre ellas. No existen cifras públicas consolidadas de usuarios o volúmenes transaccionales. Los casos de uso principales son *P2P (envíos entre usuarios)*, pagos con código *QR* en comercios y recargas de servicios. Las barreras a la adopción incluyen baja bancarización rural, desconfianza en medios digitales y costos regulatorios (IVA, tarifas). Los *drivers* de crecimiento son la penetración masiva de *smartphones* y políticas de inclusión financiera (cuentas gratuitas de dinero electrónico). (Ver Tabla N°3)

Tabla 3

Ecosistema de billeteras digitales en Ecuador

Billetera	Emisor	Casos de Uso Principales	Alcance
SIPY	Banco del Austro	Pagos QR, servicios públicos	Cuenca, nacional
PayPhone	Movilnet / Cooperativas	Recargas, P2P	Red cooperativa
Deuna	Banco Pichincha	Pagos QR, transferencias	Nacional
PeiGo	Banco Guayaquil	Pagos QR, transferencias	Guayaquil, nacional
BIMO	Banred / Pichincha	P2P, pagos móviles	Nacional
WIP	Produbanco / Guayaquil	P2P, pagos móviles	Guayaquil, internacional
Panapay	Coonecta / Cooperativas	Pagos intercooperativos	Sector cooperativo

Nota. Elaboración propia con datos de DataReportal (2025)

PROCESAMIENTO Y VISUALIZACIÓN DE DATOS

Base de datos - Personas

Para el desarrollo del presente estudio se solicitó la base de datos de personas, la cual contiene información demográfica, de comportamiento transaccional y de uso de servicios *P2P* de los usuarios de la billetera digital. Esta base constituye el insumo principal para identificar patrones de adopción, uso y activación de personas naturales.

La Tabla N°4 presenta las estadísticas descriptivas de las variables principales de la base de datos de personas:

Tabla 4

Estadística Descriptiva Base Personas

Variable	Media	Desviación estándar	Registros faltantes
Edad	34.42	11.14	13
Saldo cuenta	613.93	3,913.53	31
Ticket promedio	132,267.63	11,500.3	0
Volumen total cantidad transacciones	12.550	22.320	4,859
Volumen total monto transacciones	370,474.60	19,241,440	4,859

Nota. *Elaboración propia*

Atributos contenidos en la base personas

La base de datos incluye variables demográficas, financieras y transaccionales relevantes para el análisis de activación y uso de la billetera digital:

- cliente_nuevo: indica si el cliente es nuevo en la billetera digital.
- edad: edad del cliente.
- empleado_entidad: identifica si el cliente es colaborador de la entidad financiera.
- estado: estado del cliente (activo / inactivo).
- género: género del cliente.
- identificación: número de identificación del cliente (atributo técnico, no analítico).
- producto_cuenta: tipo de cuenta bancaria asociada.
- región_portafolio: región comercial asignada al cliente.
- saldo_cuenta: saldo disponible en la cuenta bancaria.
- segmento: segmento comercial del cliente según nivel de ingresos (masivo $\leq$ USD 1,200; masivo afluente $>$ USD 1,200 y $\leq$ USD 1,400; afluente $>$ USD 1,400).
- ticket_promedio: monto promedio de las transacciones realizadas.
- tipo_wallet: tipo de billetera utilizada.
- transacción_más_utilizada: tipo de transacción ejecutada con mayor frecuencia.
- volumen_total_cantidad_transacciones: número total de transacciones realizadas.

- volumen_total_monto_transacciones: monto total transaccionado.

Proceso de limpieza y preparación de la base de personas

La base inicial contiene aproximadamente 7,200 registros. Durante el análisis exploratorio se identificó que las variables relacionadas con volumen transaccional (monto total y cantidad de transacciones) presentaban alrededor de 4,300 valores faltantes.

Desde el contexto del negocio, estos valores no representan información errónea ni incompleta, sino que corresponden principalmente a clientes inactivos, es decir, usuarios que descargaron la aplicación, pero no realizaron transacciones. No obstante, mantener estos registros sin actividad transaccional aporta poco valor para el análisis de patrones de uso y segmentación basada en comportamiento.

Con el objetivo de fortalecer la calidad del análisis y enfocar los modelos en usuarios con señales reales de comportamiento, se decidió imputar y depurar estos registros, obteniendo una base final de aproximadamente 2,600 clientes con actividad transaccional.

Aunque esta decisión reduce significativamente el tamaño del *dataset*, permite:

- Contar con una base más balanceada entre clientes activos e inactivos.
- Mejorar la precisión y estabilidad de los modelos analíticos.

Como se observa en las **Figuras 3 y 4**, la retención de usuarios activos refleja el impacto de esta depuración en la calidad del análisis a nivel de retención e incremento de transaccionabilidad.

Figura 3

Retención de Usuarios Activos

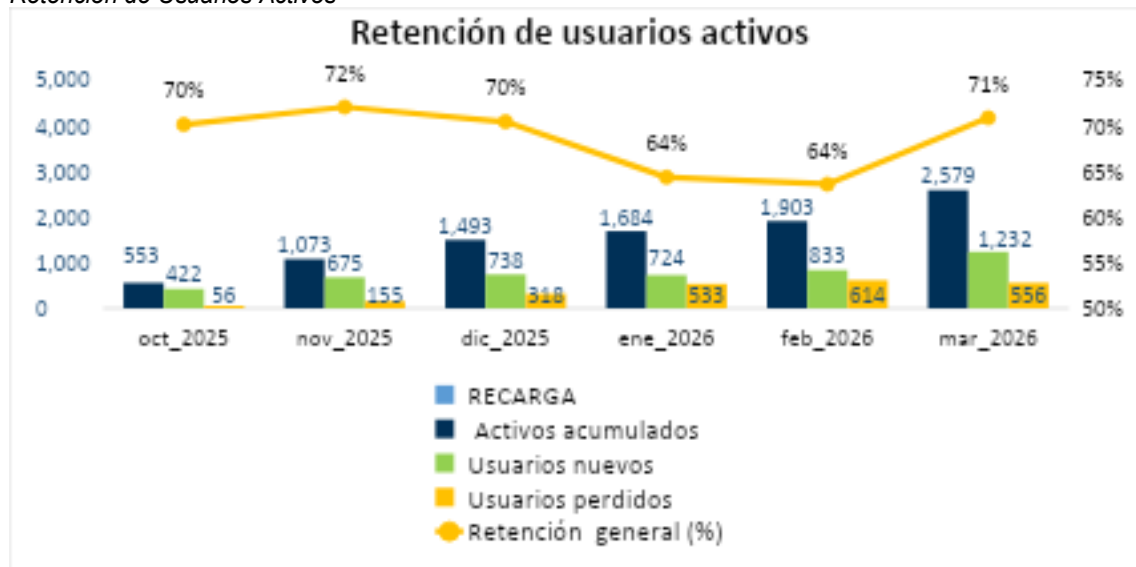
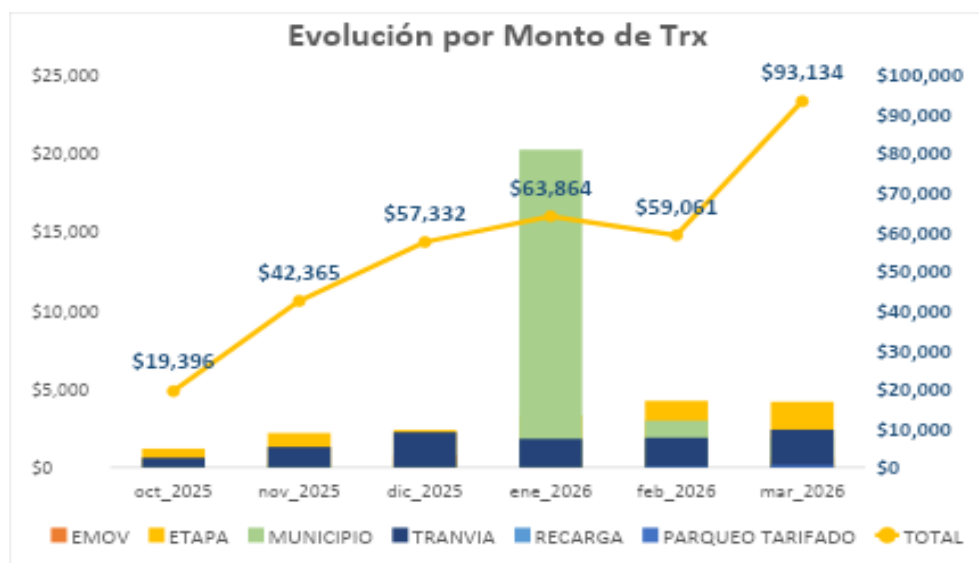


Figura 4

Evolución por Monto de Transacción



Base de datos - Comercios

Adicionalmente se solicitó la base de datos de comercios, la cual contiene información de comportamiento transaccional de los comercios de acuerdo al servicio “Pago a Comercio” o también conocido como modelo *P2B*. Esta base constituye el insumo principal para identificar patrones de adopción, uso y activación de comercios de acuerdo al número de personas realizan pagos en los mismos.

La Tabla N°5 presenta las estadísticas descriptivas de las variables principales de la base de datos de comercios:

Tabla 5

Estadística Descriptiva Comercios

Variable	Media	Desviación estándar	Registros faltantes
Saldo cuenta	874.36	5,383.09	5
Ticket promedio	5.65	27.35	0
Total monto transacciones	20.15	109.21	377
Total transacciones	1.44	9.19	377

Nota. *Elaboración propia*

Variables de la base:

- Cliente_nuevo: comercio nuevo en la billetera digital
- Empleado_entidad: comercio vinculado a empleado
- Estado: estado del comercio
- Identificación: RUC o identificación del comercio
- Producto_cuenta: tipo de cuenta asociada
- Región_portafolio: región del comercio
- Saldo_cuenta: saldo del comercio
- Segmento: segmento comercial
- Ticket_promedio: valor promedio por transacción
- Tipo_negocio: tipo de comercio
- Tipo_wallet: tipo de billetera usada
- Total_monto_transacciones: monto total procesado
- Total_transacciones: número total de transacciones

Se identificó un total de 106 campos faltantes pero al analizar este grupo de comercios se encontró que pertenecían a un grupo que se había descargado únicamente la billetera si ejecutar su primer transacción de “cobro”.

Proceso de preparación:

- Imputación de "missing values", la base quedó en 2.9K registros.
- No existen datos atípicos significativos.
- Se contaba alrededor de 500 categorías de comercios, por lo que se procedió a disminuir a 50 realizando una homologación con MAP.
- Normalización: z-transformation.
- Nominal a numéricas: dummy coding

Una vez realizada la primera depuración de datos se identificó el comportamiento de los comercios de acuerdo a su estado de activo e inactivo más cercano a la realidad del concepto, de igual manera se identificó el tipo de comercio por monto transaccionado evitando los sesgos que se generaban por incluir comercios que nunca ejecutaron una sola transacción (Ver Figuras N°4 y N°5).

Figura 5

Distribución de Comercios por Estado

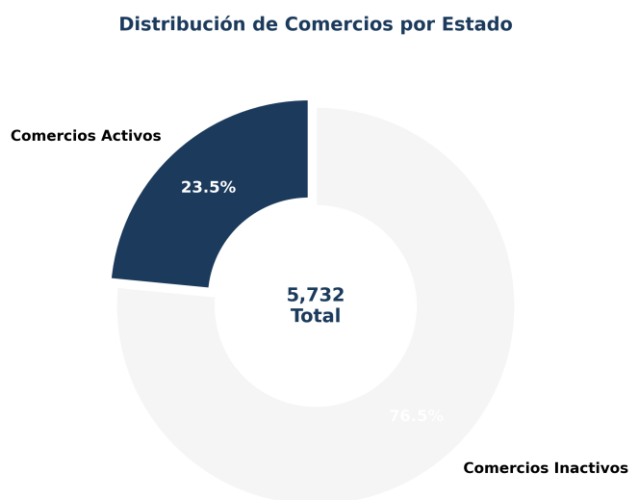
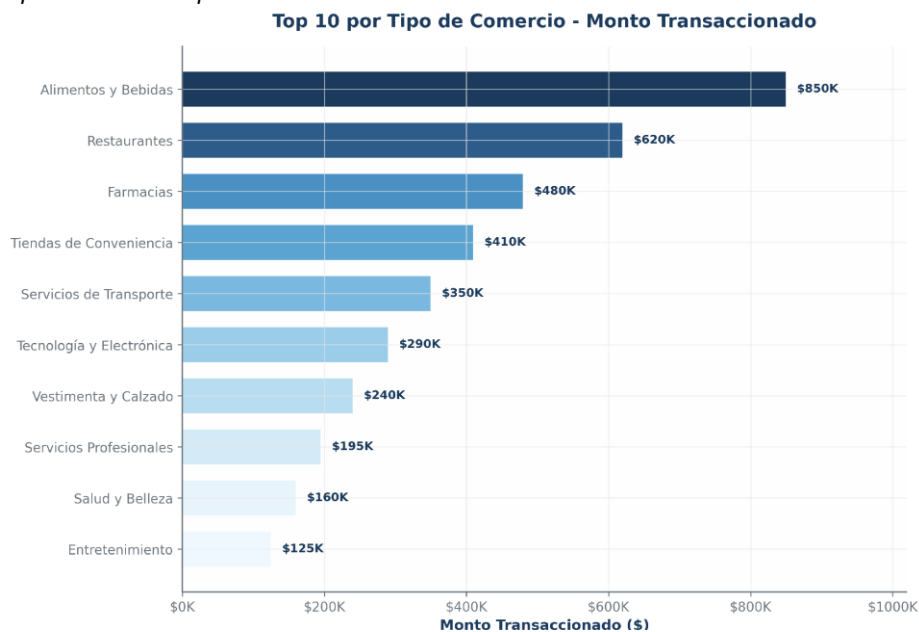


Figura 6

Tipo de Comercio por monto transaccionado



COMPETITIVIDAD Y TRANSFORMACIÓN DIGITAL

Aplicación de la Matriz PEYEA

Se aplicó la Matriz PEYEA para definir la posición estratégica de la billetera digital y priorizar las principales líneas de acción para su crecimiento, mediante el análisis estructurado del mercado, competidores y usuarios.

La metodología PEYEA permite hacer una evaluación estratégica para identificar el posicionamiento de un producto o servicio dentro del mercado en el que se desenvuelve, para evaluar las acciones estratégicas que se deben tomar para impulsar su crecimiento sostenible en el tiempo.

Matriz FODA

A través de una Matriz FODA (ver Tabla N°6) se evaluó el entorno externo e interno para un primer análisis de mercado referente a billeteras digitales (Tabla N°3):

Tabla 6

Matriz FODA

Fortalezas	Oportunidades
• Alianzas municipales (Pago de servicios) • Funcionalidad de educación financiera (comportamiento de pago) • Respaldo institucional • Equipo técnico especializado • Time to market competitivo • Alianzas grupos económicos • Funcionalidades customer centric • Acreditación en tiempo real	• Expansión Nacional • Alianzas con empresas de transporte público y privado • Integración con bancos, cooperativas y fintech • Inclusión financiera en sectores donde predomina el efectivo • Canales municipales para proyectar marca
Debilidades	Amenazas
• Intermittencia servicio • Bajo número de usuarios activos • Falta de claridad en la comunicación • Bajo número de establecimientos para pagos P2B • Dependencia de proveedores externos para biometría (onboarding) • Falta de soporte al usuario final	• Posicionamiento de otras billeteras en el mercado • Ingreso de nuevas billeteras digitales • Normativa Banco Central para controlar las billeteras digitales • Tendencia en el mercado por la interoperabilidad • Cambio de gobierno local • Presión por incentivos para adopción • Fragmentación de mercado

Nota. Elaboración propia

Calificación y ponderación de variables

En la Tabla N°7 se evidencia la matriz de calificación de cada variable en función de la importancia de cada variable (+1 a +6 para Fortalezas/Oportunidades, -1 a -6 para Debilidades/Amenazas).

Se consolidaron grupos multidisciplinarios para identificar las variables y asignar una calificación a cada una.

Tabla 7

Calificación y ponderación variables PEYEA

#	Oportunidad	Calificación	Ponderación	Puntaje ponderado
1	Consolidar transporte público y privado	6	16%	0.95
2	Integración con bancos, cooperativas y Fintech	6	42%	2.53
3	Transformar en canal de pago	6	16%	0.95
4	Canales municipales para proyectar marca	5	5%	0.26
5	Atacar segmentos más jóvenes (15 a 23)	4	11%	0.42
6	Inclusión financiera: uso de efectivo	4	11%	0.42
Puntaje Ponderado				5.53
#	Fortaleza	Calificación	Ponderación	Puntaje ponderado
1	Acreditación en tiempo real	3	3%	0.10
2	Pagos de servicios y alianza municipal	6	26%	1.50
3	Respaldo institucional - posicionamiento	5	6%	0.30
4	Time to market competitivo	4	16%	0.60
5	Entidad financiera + empresas del corporativo	5	13%	0.60
6	Funcionalidades	1	16%	0.20
7	Educación financiera	6	3%	0.20
8	Nuevo canal para clientes de la entidad financiera	6	10%	0.60
9	Afinidad con migrantes	5	3%	0.20
10	Volumen de clientes existentes	4	3%	0.10
Puntaje Ponderado				4.50
#	Amenaza	Calificación	Ponderación	Puntaje ponderado
1	Posicionamiento billeteras del mercado	-6	24%	-1.43

2	Normativa BCE o SB que podrían sacar contra billeteras	-3	19%	-1.43
3	Estandarización del mercado - Interoperabilidad	-6	5%	-0.57
4	Ingreso de nuevas billeteras: Claro, Banco Pacifico, etc.	-2	14%	-0.29
5	Resistencia de los comercios a la afiliación	-2	5%	-0.29
6	Cambio del gobierno local	-6	10%	-0.10
7	Presión por incentivos para adopción	-6	10%	-0.57
8	Fragmentación del mercado	-3	14%	-0.57
Puntaje Ponderado				-5.24
#	Debilidad	Calificación	Ponderación	Puntaje ponderado
1	Estructura Legal	-5	10%	-0.50
2	Baja retención	-6	10%	-0.60
3	Intermitencia en el Servicio	-6	20%	-1.20
4	Definición modelo de presupuesto	-3	15%	-0.45
5	Falta de claridad de comunicación y alcance	-4	15%	-0.60
6	Pocos establecimientos afiliados y activar	-6	10%	-0.60
7	Inexistencia de herramientas para producto Digital	-3	5%	-0.15
8	Completar funcionalidades de mercado	-3	5%	-0.15
9	Área de soporte	-6	10%	-0.60
Puntaje Ponderado				-4.85

Nota. Elaboración propia

Ubicación en el cuadrante PEYEA

Una vez obtenidos los puntajes ponderados por cada frente se procede a calcular el valor final para ubicar de acuerdo a las coordenadas respectivas e identificar el cuadrante en el que la billetera digital se ubica (Ver Tabla N°8).

Tabla 8

Cálculo de Coordenadas Matriz PEYEA

Variable	Valor
Fortaleza Financiera + Ventaja Competitiva (X ponderado)	-0.37

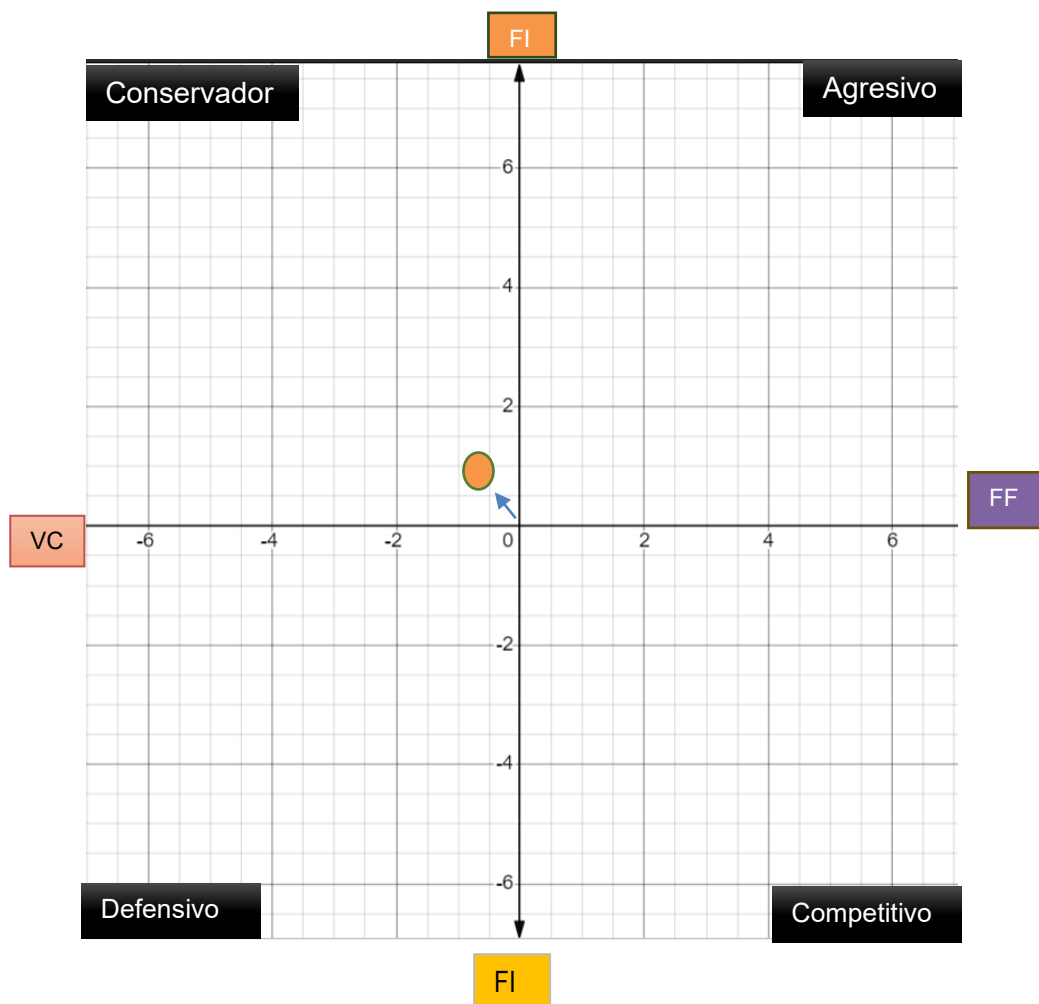
Estabilidad del Entorno + Fuerza de la Industria (Y ponderado)	0.29
--	------

Nota. Elaboración propia

Posterior al cálculo, como se observa en la Figura N°8 la intersección de las coordenadas ubicaron a la billetera digital en el cuadrante “Conservador”

Figura 7

Matriz PEYEA



Definición de estrategias

El Taller Estratégico ha posicionado a la organización en el Cuadrante Conservador de la Matriz de Posicionamiento Estratégico. Este posicionamiento indica que, si bien el entorno de mercado presenta oportunidades favorables, la billetera digital aún no dispone de la fortaleza interna suficiente para escalar agresivamente.

"El paso lógico no es cambiar la estrategia, sino fortalecer las capacidades internas para poder migrar hacia un cuadrante competitivo o agresivo."

Diagnóstico del cuadrante conservador:

El entorno es favorable (hay oportunidad) pero la billetera digital todavía no tiene suficiente fortaleza interna para escalar agresivamente.

Implicación: Hay oportunidad, pero la billetera digital todavía no está lista para aprovecharla al máximo. La prioridad estratégica debe ser fortalecerse antes de crecer.

Objetivo estratégico: Fortalecerse antes de crecer

Estrategias clave:

- 1. Penetración del mercado (controlada): Activar clientes actuales sin grandes inversiones; Campañas focalizadas (no masivas); Crecer sin sobreextenderse.
- 2. Desarrollo del producto: Mejorar funcionalidades básicas; Eliminar fricciones de uso; Garantizar estabilidad tecnológica.
- 3. Fortalecimiento interno: Optimizar procesos internos; Mejorar soporte operativo; Alinear equipos y talento.

Prioridades a considerar:

- 1. Estabilizar el producto: 'Sin un producto confiable, no hay adopción sostenible.' Mejorar la disponibilidad (intermitencia actual); Mejorar UX en puntos críticos de onboarding.
- 2. Activación y comercios: 'La billetera solo existe si se usa.' Afiliación real de comercios; Uso real - frecuencia de uso por parte de usuarios; Engagement e incentivos para el uso recurrente.
- 3. Claridad de propuesta de valor: ¿Por qué usar esta billetera digital? Propuesta de valor sólida y diferenciada; Factores diferenciadores frente a la competencia.
- 4. Comunicación efectiva: 'Si el cliente no entiende el valor, no adopta.' Segmentar comunicación por perfiles de usuario; Educación al Cliente sobre propuesta de valor; Mensajes claros y accionables.

GESTIÓN Y DISEÑO DE PROYECTOS BIG DATA

Objetivo general del proyecto

Incrementar las tasas de activación mediante la aplicación de segmentación avanzada y modelos predictivos, que permitan priorizar los esfuerzos comerciales hacia personas y comercios con mayor probabilidad de uso transaccional y generación de valor.

Objetivos del negocio

- Generar la adquisición de 50 mil usuarios activos (30%) durante el primer año.
- Generar la adquisición de 12 mil comercios activos (30%) durante el primer año.
- Disminuir el costo de adquisición de usuarios en un 5%, mediante una focalización más eficiente de las campañas comerciales.

Objetivos de la analítica avanzada

- Segmentar personas y comercios que hayan transaccionado al menos una vez en la billetera digital, identificando patrones de comportamiento y uso.
- Desarrollar un modelo de predicción de usuarios activos, que permita identificar potenciales usuarios con alta probabilidad de activación, tanto dentro del portafolio actual de la billetera digital (usuarios inactivos) como en la base de clientes de la entidad financiera que aún no utilizan la billetera.

Criterios de éxito

Incremento proyectado en las tasas de activación:

- Personas: del 18% al 40%.
- Comercios: del 23% al 30%.

Obtención de modelos analíticos con métricas robustas y consistentes, tales como:

- Silhouette y Davies-Bouldin para la evaluación de clustering.
- Accuracy superior al 80% en los modelos de clasificación.

Modelado

Modelo no supervisado: *Clustering*

Se aplicó *K-Means* como técnica principal de *clustering*, debido a su eficiencia y facilidad de interpretación en contextos de negocio. Para la selección del número óptimo de *clusters* se utilizó el índice de Davies-Bouldin, ejecutando el modelo en un *loop* de parámetros con valores de *k* entre 2 y 8.

Evaluación del modelo

El índice de Davies-Bouldin indicó que el número óptimo de *clusters* para personas es $k = 4$, al presentar la menor relación entre dispersión *intra-cluster* y separación *inter-cluster*.

Para comercios, el valor óptimo fue $k = 3$, logrando una segmentación clara y balanceada del portafolio activo.

Variables utilizadas

• Personas: edad, volumen total de transacciones, monto total transaccionado, *ticket* promedio, *log(saldo)*, género, segmento y tipo de transacción más utilizada. • Comercios: total de transacciones, monto total, *ticket* promedio y *log(saldo)*. • Las variables categóricas (segmento, tipo de negocio) se incorporaron únicamente para el perfilado posterior, evitando sesgos en la distancia euclidiana.

Resultados del clustering - Personas ($k = 4$)

Cluster 0 - 'Usuarios ligeros / cotidianos' ($P2P$ + Transporte)

Perfil numérico: Frecuencia: -0.21 (baja), monto total: -0.22 (bajo), *ticket*: -0.13 (bajo), Edad ~ promedio (-0.016).

Canal dominante:

- $P2P$ 51%
- Transporte 38% • Municipio ~2.7%

Análisis: usuarios con pagos pequeños, más de $P2P$ y algo de transporte.

Campañas recomendadas:

- "Reto 5 transacciones" (activar hábito)
- *Micro-cashback* por recurrencia (no por monto)
- Activaciones en comercios cotidianos (cafeterías, minimarkets)

Cluster 2 - 'Transporte intensivo / alta frecuencia - bajo *ticket*'

Perfil numérico: Frecuencia: +3.53 (altísima), monto total: +0.50 (moderado), *ticket*: -0.32 (bajo), edad: -0.097 (ligeramente más joven)

Canal dominante:

- Transporte 84% (clarísimo)

Análisis: usuarios recurrentes, muchas transacciones pequeñas (ideal 'uso hábito').

Campañas recomendadas:

- Fidelización por frecuencia (niveles, insignias, beneficios)
- *Referidos*: este grupo es perfecto para traer nuevos usuarios
- *Cross-sell* a "Municipio" (ya confían y usan)

Cluster 1 - 'Municipio *power users* / alto volumen'

Perfil numérico: Frecuencia: +0.67 (sobre el promedio), monto total: +4.03 (muy alto) , *ticket*: +1.37 (alto) , *log_saldo*: +0.43 (ligeramente alto)

Canal dominante:

- Municipio 49%
- *P2P* 36%
- Transporte 7.5%

Análisis: usuarios que mueven dinero fuerte y usan mucho el caso municipal (el convenio estrella).

Campañas recomendadas:

- Beneficios por trámites municipales recurrentes
- "Paquetes" de pagos municipales (recordatorios, autogestión)
- *Upsell* a comercios: "paga con la billetera digital en comercios aliados" (alto valor)

Cluster 3 - 'Alto *ticket* ocasional / nicho *premium*'

Perfil numérico: • Frecuencia: -0.49 (baja) • Monto total: +0.95 (alto) • *Ticket*: +6.42 (muy alto) • Edad: +0.51 (más mayores)

Canal dominante: • Municipio 34% • *P2P* 40% • "Otros" 15.6%

Análisis: usuarios que transaccionan poco pero cuando lo hacen, hacen pagos grandes.

Campañas recomendadas:

- "2da transacción" (incentivo a repetir)
- Mensaje *premium*: seguridad, confiabilidad, rapidez

- Productos de alto valor / comercios ancla (salud, educación, etc.)

Clustering - Comercios ($k = 3$)

Se realizó un primer ejercicio con la base considerando comercios inactivos y activos, pero existe muchísimo sesgo hacia los inactivos y esto guarda relación con el bajo porcentaje de comercios activos que se tiene (~18%). Para *clustering* se decidió solo filtrar los activos, considerando solo una base de 400 comercios, pero así permitió identificar y balancear de mejor manera los *clusters* identificando el arquetipo de comercios que están transaccionando con la billetera digital.

De acuerdo al índice de Davies-Bouldin, el número ideal de *clusters* es 3.

Resultados *clustering* (K-Means, $k = 3$)

Para segmentar a los comercios por su comportamiento transaccional en la billetera digital, se aplicó un modelo *K-Means* con $k = 3$, usando exclusivamente las variables numéricas: *total_transacciones*, *total_monto_transacciones*, *ticket_promedio* y *Log_saldo*.

Posteriormente, los *clusters* se perfilaron mediante *tipo_negocio* y segmento usando un *join*, garantizando que estas variables categóricas no distorsionen las distancias del algoritmo.

Los *clusters* representan tres niveles claros de desempeño transaccional: Bajo, Medio y Alto Valor.

Cluster 0 - Segmento Base / Bajo Dinamismo (Es el *clúster* más grande y representa la mayoría del portafolio)

Perfil general: • La gran mayoría de variables estandarizadas aparecen negativas (entre -0.10 y -0.35). • Representa comercios con menor actividad relativa, montos bajos y *tickets* bajos. • Incluye una alta variedad de giros del segmento MASIVO.

Principales giros presentes:

- Autoservicios/Bodegas MASIVO
- Bazar/Librería MASIVO
- Panadería/Pastelería MASIVO
- Restaurantes MASIVO
- Servicio Técnico MASIVO
- Spa/Belleza MASIVO • Venta de Ropa/Calzado MASIVO

Análisis: Este *clúster* agrupa a los comercios que realizan transacciones pequeñas, con frecuencia baja a media, y montos inferiores al promedio. Es el grupo más estable en tamaño, pero con menor dinamismo.

Cluster 1 - Segmento Intermedio / Dinamismo Promedio

Perfil general esperado: • Valores cercanos a 0 en las estandarizadas (desempeño promedio). • Combina comercios con actividad moderada, *tickets* estables y montos normales.

Lectura de negocio: Este *clúster* suele representar comercios que ya utilizan la billetera digital regularmente, pero aún no alcanzan volúmenes altos ni *tickets* diferenciados.

Clúster 2 - Segmento de Alto Valor / Alto Ticket

Los valores representados en la Tabla N°9 muestran un desempeño positivo:

Tabla 9

Características Segmento Alto Valor

Variable	Valor
avg(total_transacciones)	-0.13 (frecuencia normal o algo baja)
avg(total_monto_transacciones)	1.34 (MUCHO MAYOR al promedio)
avg(ticket_promedio)	2.58 (extremadamente alto)
avg(log_saldo)	0.75 (saldo más alto)

Nota. *Elaboración propia*

Análisis: pagos de alto monto, tickets altos y comercios de gran valor, incluso si la frecuencia no es tan elevada.

Giros presentes:

- Venta de Ropa / Calzado (MASIVO)

Este *clúster* representa comercios que generan alto volumen monetario, con tickets significativamente superiores a la media. Son comercios con alto valor para la billetera digital, aunque su frecuencia no sea extrema.

Modelo supervisado: Predicción de activación

Se evaluaron distintos modelos supervisados con el objetivo de predecir activación y reactivación de usuarios:

- Árbol de Decisión
- Random Forest
- PRISM
- Redes Neuronales

El "label" utilizado fue estado del usuario (activo / inactivo), ajustando pesos para priorizar la clase de interés según el objetivo del modelo.

Árbol de Decisión y Random Forest

Inicialmente, el Árbol de Decisión alcanzó un *accuracy* del 59%, el cual se incrementó a 71% tras ponderar la clase activa. Posteriormente, Random Forest permitió mejorar estabilidad y desempeño:

- Accuracy: 71.30% $\pm$ 3.57%
- Recall (activo): 80.09% $\pm$ 3.54%
- Precision (activo): 75.23% $\pm$ 2.86%

Este modelo es adecuado para campañas de activación, al priorizar la detección de usuarios activos potenciales.

PRISM

El modelo PRISM alcanzó un *accuracy* elevado (86.44%). Sin embargo presentó un recall extremadamente bajo para inactivos (5.69%), debido al fuerte desbalance del dataset.

Esto evidencia que, aunque el *accuracy* es alto, el modelo no es adecuado para identificar inactivos, por lo que su uso se limita a análisis exploratorio y no operativo.

Red Neuronal

La red neuronal fue aplicada específicamente para el objetivo de reactivación de usuarios inactivos, ajustando pesos hacia dicha clase.

Configuración final: 2 capas ocultas (20 y 30 neuronas), 1,600 ciclos de entrenamiento; Learning rate = 0.01

Validación cruzada

Resultados:

- Accuracy: 69.47% $\pm$ 2.62%
- Recall (inactivo): 75.61% $\pm$ 7.37%
- Precision (inactivo): 75.24% $\pm$ 5.12%
- AUC: 0.762 $\pm$ 0.039

Este modelo logra un equilibrio adecuado entre cobertura y eficiencia, siendo el más alineado al objetivo de reactivación.

Evaluación

Todos los modelos fueron evaluados mediante cross-validation, garantizando estabilidad de resultados y evitando sobreajuste.

- Random Forest: mejor desempeño para identificar activos (activación).
- Red Neuronal: mejor modelo para reactivar inactivos.
- PRISM: alto accuracy engañoso por desbalance, bajo valor operativo.

Variables críticas identificadas

Las siguientes variables explican la mayor parte del comportamiento de activación y valor:

- Frecuencia de transacciones
- Monto total transaccionado
- Ticket promedio
- Tipo de transacción
- Segmento del cliente / comercio

Resultados esperados e impacto estratégico

Impacto en el negocio

Focalización precisa de la fuerza comercial hacia segmentos de alto potencial. Incremento proyectado de activación:

- Personas: 18% → 40%
- Comercios: 23% → 30%
- Reducción del costo de adquisición mediante campañas dirigidas.
- Mayor retención de usuarios y comercios de alto valor.

Impacto en analítica avanzada

- Segmentación reutilizable y dinámica.
- Modelos predictivos alineados a objetivos reales de negocio.
- Base sólida para futuros modelos de cross-selling, up-selling y propensión a productos financieros.

ECONOMÍA Y FINANZAS PARA LA GESTIÓN DE PROYECTOS

Contexto de rentabilidad

Uno de los retos más significativos del mercado es encontrar la rentabilidad de una billetera digital a partir de la gran inversión que estas representan para alcanzar una posición estratégica en el share de mercado. Muchas iniciaron con un cobro 0 por transacción. Sin embargo, los costos por publicidad y la presión por incentivos para impulsar la activación de usuarios han impulsado el cobro de una tarifa MDR a ciertos segmentos de micro comercios.

Proyección financiera

La Tabla N°10 presenta la proyección financiera del proyecto a 5 años, considerando la estructura de *clústers* identificada y las tarifas MDR diferenciadas:

Tabla 10

Proyección Financiera Billetera Digital

Concepto	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5
Adquisición usuarios		166,000	250,000	350,000	450,000	500,000
Adquisición comercios		40,000	80,000	120,000	160,000	200,000
% activación usuarios		30%	35%	40%	45%	45%
% activación comercios		30%	35%	40%	45%	45%
Real activados usuarios		49,800	87,500	140,000	202,500	225,000
Real activados comercios		12,000	28,000	48,000	72,000	90,000
Total Ingresos MDR		\$42,120	\$127,680	\$279,360	\$533,520	\$810,000
Total Cross-selling			\$1,046,000	\$1,528,000	\$2,310,000	\$3,246,000
Total Egresos		\$792,000	\$934,000	\$1,357,000	\$1,537,000	\$1,820,000
Inversión inicial	\$2,000,000					
Flujo Neto	-\$2,000,000	-\$749,880	\$239,680	\$450,360	\$1,306,520	\$2,236,000

Nota. *Elaboración propia*

Indicadores financieros VAN y TIR

- VAN (Valor Actual Neto): \$589,766.81
- TIR (Tasa Interna de Retorno): 76%

Análisis: El proyecto presenta un VAN positivo y una TIR del 76%, lo que indica que la inversión es financieramente viable. Los flujos negativos en el año 0 y año 1 corresponden a la inversión inicial y los costos de activación, mientras que desde el año 2 el proyecto genera flujos positivos crecientes gracias a la monetización por MDR y cross-selling.

Estructura de tarifas *MDR*

Tomando los supuestos de las fuentes consultadas en el punto *Big Data Solutions*, se plantea que la billetera digital, en función de aplicar las estrategias definidas previamente, pueda acceder a parte de ese share de mercado e impulsar una mayor adquisición y activación de usuarios. Además, a través de los clústers identificados, cobrar una tarifa diferenciada considerando variables como el nivel de ventas, tamaño, entre otras (Ver Tabla N°11)

Tabla 11

Tarifa por Segmento de Comercio

Segmento de comercios	MDR x TRX
Cluster segmento Base	0%
Cluster segmento Afluente	2%
Cluster Alto Valor	3%

Nota. *Elaboración propia*

CONCLUSIÓN

La aplicación de la metodología *CRISP-DM* al caso de estudio de la billetera digital demuestra que la analítica avanzada constituye un habilitador clave para la transformación digital y comercial de la institución financiera. A través de un enfoque estructurado, fue posible convertir datos transaccionales y de comportamiento en conocimiento accionable, alineando de manera efectiva los objetivos del negocio con las decisiones analíticas.

La combinación de modelos no supervisados (*clustering*) y modelos supervisados permitió no solo comprender el comportamiento actual de personas y comercios, sino también anticipar qué perfiles presentan mayor probabilidad de activación y reactivación. El *clustering* evidenció arquetipos claramente diferenciados de usuarios y comercios, facilitando la definición de estrategias segmentadas de adquisición, fidelización y monetización. Por su parte, los modelos predictivos —en particular *Random Forest* y Redes Neuronales— mostraron un desempeño adecuado para priorizar esfuerzos comerciales, destacándose la importancia de métricas por clase (*precision*, *recall* y *AUC*) frente a la *accuracy* global en contextos desbalanceados.

En conjunto, los resultados confirman que una estrategia basada en analítica avanzada permite optimizar la focalización de la fuerza comercial, reducir el costo de adquisición y elevar las tasas de activación efectiva, sentando las bases para un crecimiento sostenible. Este trabajo evidencia cómo la integración entre datos, negocio y estrategia convierte a la analítica en un pilar fundamental para la toma de decisiones y la generación de valor en el ecosistema digital de la institución financiera.

Recomendaciones estratégicas

- No escalar masivamente hasta consolidar la base operativa y tecnológica.
- Priorizar la estabilidad del producto como condición necesaria para cualquier crecimiento posterior.
- Focalizar recursos en activación real (usuarios y comercios) sobre captación masiva de nuevos usuarios.
- Fortalecer la propuesta de valor con mensajes claros y diferenciados frente a la competencia.
- Alinear operación con la red de pagos como base transversal de gobernanza y cumplimiento.

REFERENCIAS BIBLIOGRÁFICAS

Micus, M., Schramm, M., Boehm, M., & Krcmar, H. (2023). How big data analytics enables innovation and value creation: A systematic literature review. *Journal of Business Research*, 154.

Woon, W. L., Jiyeong, L., & Minjung, K. (2025). Sustainable product lifecycle management through data analytics: A systematic review. *Journal of Cleaner Production*, 435.

Zhuanghao, Z., et al. (2025). Machine learning and deep learning applications in customer satisfaction and product design: A comprehensive review. *Computers & Industrial Engineering*, 187.

Juarez-Alvarez, C. R., Merino-Lazo-De-La-Vega, V. P., Villanueva-Paredes, G. X., & Esquicha-Tejada, J. D. (2025). Adoption of digital wallets and their impact on the sales of micro-entrepreneurs. *Polish Journal of Management Studies*, 31(1).

Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, 281(3).

ARCOTEL. (2025). *Estadísticas de telecomunicaciones: Conexiones Internet fijo*. Agencia de Regulación y Control de Telecomunicaciones. Recuperado de <https://www.ecuadorencifras.gob.echttps://www.arcotel.gob.ec>

Cámara Ecuatoriana de Comercio Electrónico [CECE]. (2024). *Estudio de transacciones no presenciales en Ecuador — VII medición*. Recuperado de <https://www.ecuadorencifras.gob.echttps://www.cece.org.ec>

DataReportal. (2025). *Digital 2025: Ecuador*. Recuperado de <https://www.ecuadorencifras.gob.echttps://datareportal.com/reports/digital-2025-ecuador>

INEC. (2025). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU): Indicadores de conectividad digital*. Instituto Nacional de Estadística y Censos del Ecuador. Recuperado de <https://www.ecuadorencifras.gob.ec>

Banco Central del Ecuador. (2025). Resolución Administrativa NRO. BCE-GG-008-2025. Norma que regula los medios y sistemas de pago en Ecuador y las actividades Fintech de sus partícipes. Recuperado de https://www.bce.fin.ec/storage/BCE/sistemas_auxiliares/BCE-GG-008-2025.pdf

Cámara Ecuatoriana de Comercio Electrónico [CECE]. (2024). Estudio de transacciones no presenciales en Ecuador — VII medición. Recuperado de <https://www.cece.org.ecCredicorp>.
(2025). Índice de Inclusión Financiera (IFF) 2025. Grupo Credicorp.

Payment & Commerce Market Intelligence [PCMI]. (2024). *Ecuador e-commerce landscape 2024*. Recuperado de <https://www.pcmi.com>

ANEXOS

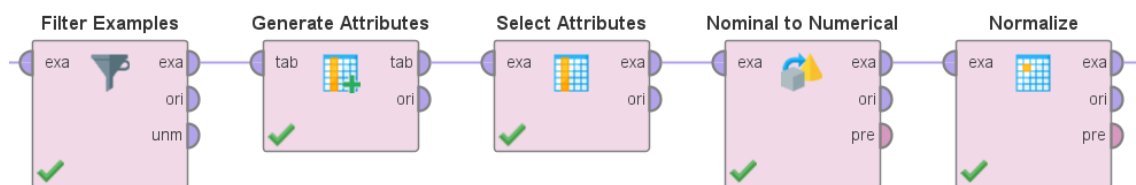
Anexo 1. Capturas del tratamiento de datos

Base sin tratamiento de datos

Name	Type	Missing	Statistics			Filter (12 / 12 attributes)	Search for Attributes
id	Integer	0	Min	1	Max	Average	
identificacion	Integer	0	Min	1	Max	Average	
genero	Nominal	13	Least	Femenino (2995)	Most	Masculino (4553)	Values
edad	Integer	13	Min	18	Max	126	Average
empleado_ba	Nominal	0	Least	SI (1096)	Most	NO (6465)	Values
cliente_nuevo_sipy	Nominal	0	Least	SI (539)	Most	NO (7022)	Values
segmento	Nominal	16	Least	PYME PLUS (2)	Most	MASIVO (6705)	Values
saldo_cuenta	Real	31	Min	0	Max	159139.990	Average
volumen_total_cantidad_transa...	Integer	4859	Min	1	Max	261	Average
transaccion_mas_utilizada	Nominal	0	Least	Parqueo Tarifado (2)	Most	Sin transacciones (4859)	Values
volumen_total_monto_transacc...	Real	4860	Min	0.020	Max	1000000000	Average
tiket promedio	Real	0	Min	0	Max	1000000000	Average
estado	Nominal	0	Least	activo (1688)	Most	inactivo (5873)	Values

Filter examples: existían datos atípicos en el volumen total y en la edad

tiket promedio	is not missing		
segmento	is not missing		
edad	≤	80	
saldo_cuenta	is not missing		
v_total_monto_transacciones	is not missing		
v_total_monto_transacciones	≤	2000	



- **En Generate Attribute** : se aplicó la siguiente formula para que no exista sesgo en los saldos $\log_saldo = \log(1 + saldo_cuenta)$ (reduce sesgo de saldos)
- **Select Attributes** : se seleccionó de la base las variables más relevantes
- **Nominal to numerical**: dummy coding
- **Normalize**: z-transformation

Anexo 2. Índice de Davies-Bouldin - Selección de *clústers*

Con *loop parameters* se corrió 8 veces el modelo de *clustering* para identificar el número ideal de *clústers* con el índice de Davies-Bouldin.

Loop Parameters (7 rows, 3 columns)

iteration	Clusteri...	Dav... ↓
3	4	-0.954
5	6	-1.115
6	7	-1.125
7	8	-1.134
4	5	-1.156
1	2	-1.325
2	3	-1.331

Nota: De acuerdo al índice de Davies-Bouldin, el número de *clústers* ideal son 4 para personas y 3 para comercios, por lo que se procede a identificar los mismos.

Anexo 3. Tabla de centroides - Personas

La siguiente tabla presenta los centroides de los 4 *clústers* identificados para la base de personas:

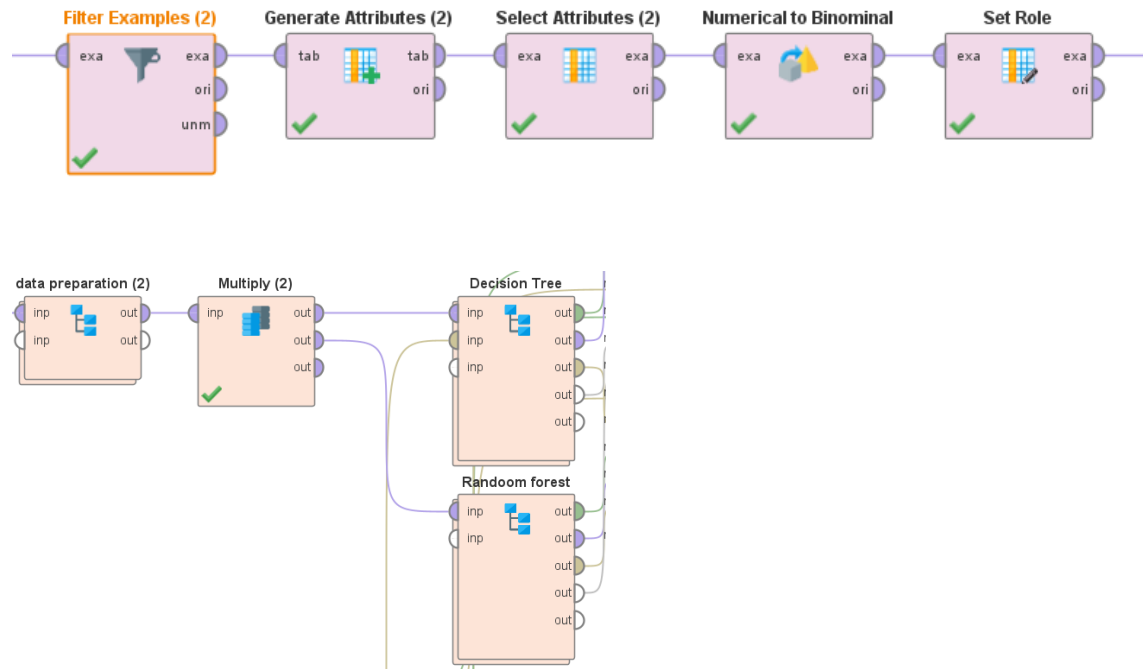
Atributo	Cluster_0	Cluster_1	Cluster_2	Cluster_3
Edad	-0.016	0.319	-0.097	0.507
volumen_total_cantidad_transacciones	-0.211	0.670	3.528	-0.489
volumen_total_monto_transacciones	-0.220	4.035	0.496	0.952
ticket promedio	-0.131	1.365	-0.322	6.416
Log_Saldo	-0.011	0.433	-0.114	-0.184
genero = Femenino	0.411	0.443	0.460	0.250
genero = Masculino	0.589	0.557	0.540	0.750
segmento = MASIVO	0.904	0.906	0.929	0.813
segmento = AFLUENTE MASIVO	0.076	0.085	0.063	0.125
segmento = AFLUENTE	0.019	0.009	0.008	0.031
transaccion_mas_utilizada = P2P	0.512	0.358	0.151	0.406
transaccion_mas_utilizada = Transporte	0.381	0.075	0.841	0.031
transaccion_mas_utilizada = Municipio	0.027	0.491	0.000	0.344

Nota: Los valores presentados corresponden a los centroides estandarizados de cada *cluster* identificado mediante el algoritmo *K-Means* con $k = 4$. Las variables *ticket* promedio, *Log_Saldo* y transacción más utilizada (*P2P*) son anglicismos técnicos utilizados en la nomenclatura original del modelo. Los valores negativos indican posiciones por debajo de la media global, mientras que los positivos reflejan posiciones por encima. Elaboración propia con datos de la billetera digital

Anexo 4. Modelos predictivos

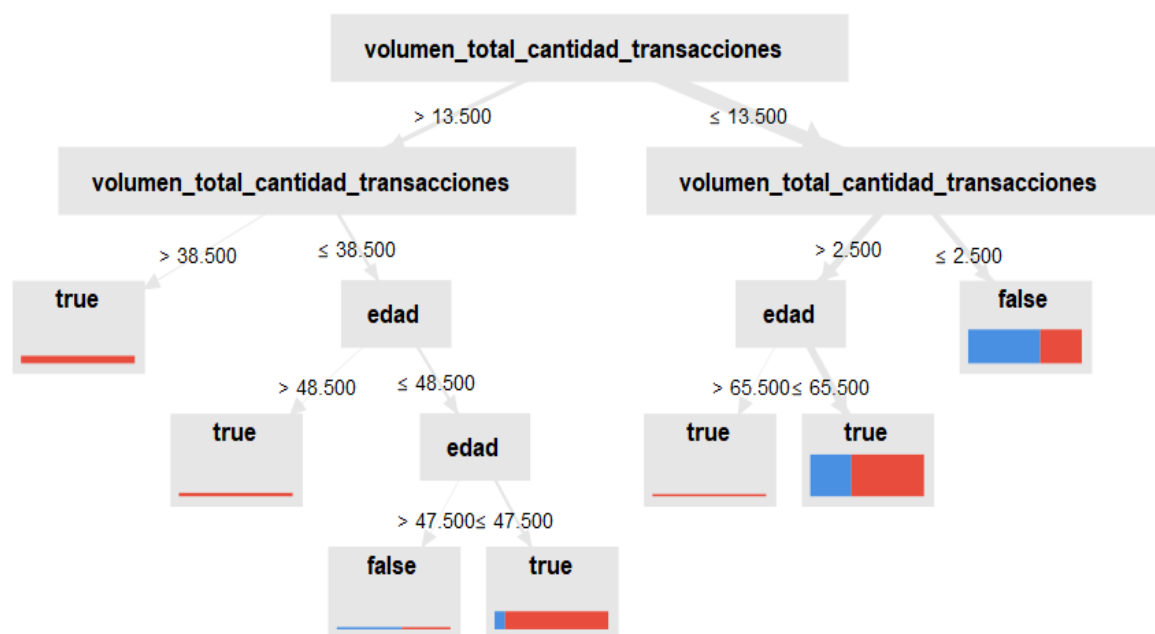
Decision Tree

Decison Tree



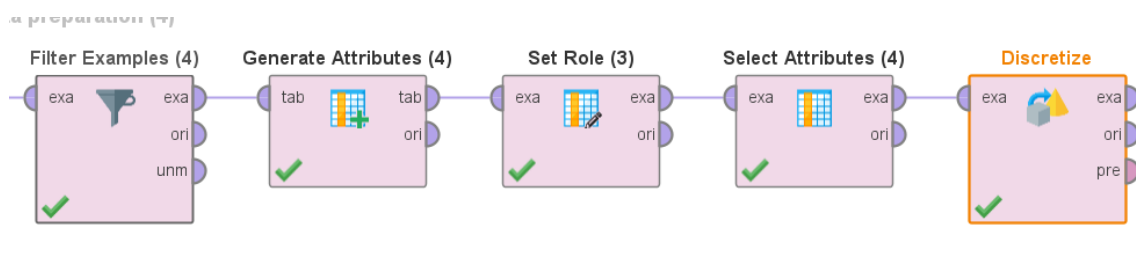
Nota: En una primera instancia se obtuvo un accuracy del 59% en el performance, pero se identificó que existen muchos inactivos. Se dio mayor peso a los usuarios activos con el operador generate attribute. Se pasó de 59% a un 71% considerado aceptable.

Random Forest:



Nota: El modelo de clasificación (árbol/flujo actual) logra un accuracy promedio de 71.30% ($\pm 3.57\%$), lo que indica un desempeño estable entre los folds de la validación cruzada. Tomando como clase positiva 'true' (equivalente a activo), el modelo obtiene:

Modelo W-Prism:



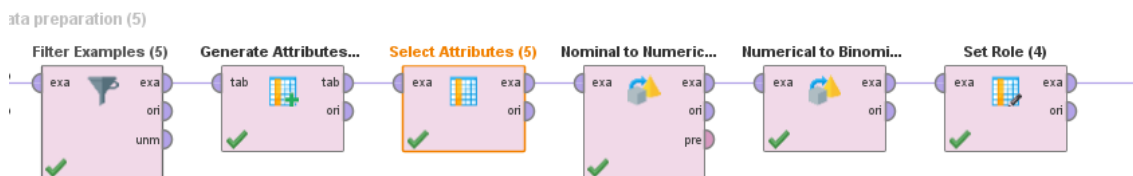
Nota: En los primeros intentos se obtuvo un accuracy de 62%, pero para mejorar el modelo se hizo un balance entre las variables. Se colocó el atributo peso como weight y estado como label para balancear el peso hacia el estado, y mejoró hasta un 86%.

Matriz de confusión PRISM:

Criterion	☒ Table View ☐ Plot View		
accuracy	accuracy: 86.44% +/- 1.17% (micro average: 86.44%)		
precision			
recall			
AUC (optimistic)			
AUC			
AUC (pessimistic)			
	true activo	true inactivo	class precision
pred. activo	6452	945	87.22%
pred. inactivo	76	57	42.86%
class recall	98.84%	5.69%	

Red Neuronal:

Configuración inicial

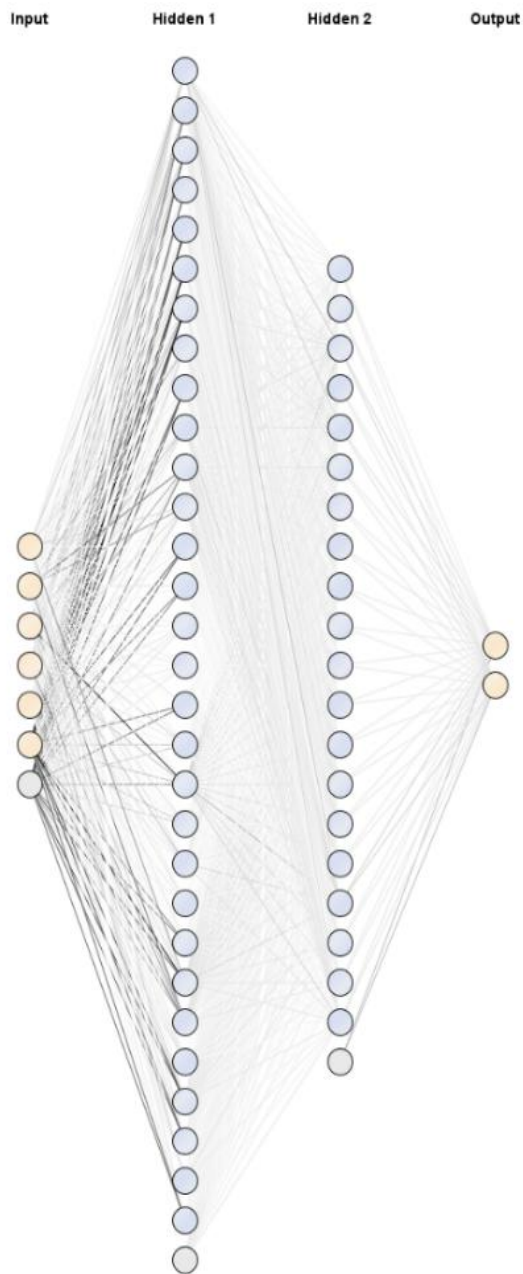


Nota: Dos capas: 20 y 10 respectivamente, Training cycles: 2,000, Learning rate: 0.005,
Resultado: accuracy de 61%

Matriz de Confusión(ajustada)

accuracy: 69.47% +/- 2.62% (micro average: 69.48%)			
	true false	true true	class precision
pred. false	596	398	59.96%
pred. true	406	1234	75.24%
class recall	59.48%	75.61%	

Nota: Dos capas ocultas: 20 y 30 respectivamente, training cycles: 1,600, learning rate: 0.01



Anexo 8. Proceso global en RapidMiner

