

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración Pública

**CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y
ANÁLISIS DE DATOS EN ORGANIZACIONES**

TRABAJO FINAL INTEGRADOR

La digitalización bancaria como variable relevante para
la gestión de deuda y segmentación de clientes morosos

Implementación de modelos de Machine Learning

AUTOR: MARIANO NAHUEL ESMAIL

TUTORA: MG. NATALIA SALABERRY

OCTUBRE 2023

Resumen

La evolución de las tecnologías y la era del Big Data han revolucionado la forma en que las organizaciones operan, especialmente en el ámbito bancario. La disponibilidad sin precedentes de volúmenes masivos de datos ha presentado tanto desafíos como oportunidades para las instituciones financieras, que buscan mejorar la toma de decisiones informadas y maximizar su eficiencia operativa. En la actualidad, las instituciones bancarias enfrentan el desafío de gestionar la morosidad de manera efectiva, buscando recuperar deudas impagas de manera eficiente y minimizando costos.

Tradicionalmente, la segmentación de la cartera morosa se ha basado en criterios como ingresos, riesgo crediticio y variables demográficas. Sin embargo, la creciente digitalización ha proporcionado una nueva dimensión de información valiosa: el nivel de interacción de los clientes con los canales digitales del banco, como la aplicación móvil, *OnlineBanking*, *Mails* y cajeros automáticos. Esta información puede ser crucial para comprender los patrones de comportamiento de los clientes morosos y personalizar estrategias de recuperación.

En este contexto dinámico, el presente trabajo tiene como objetivo principal optimizar la gestión de la morosidad en una entidad bancaria local¹ mediante la incorporación de una dimensión clave: el nivel de digitalización de los clientes. Para alcanzar a este objetivo, se proponen tres objetivos interconectados. En primer lugar, se busca identificar patrones en el nivel de digitalización de los clientes y su relación con el comportamiento de morosidad. A partir de analizar datos de la organización bajo análisis se determinará cómo las interacciones digitales pueden indicar tendencias de pago y generar patrones de comportamiento en la cartera de clientes morosos.

En segundo lugar, se propone aplicar técnicas de *Clustering* para segmentar a los clientes morosos en grupos más específicos, considerando tanto las variables tradicionales, así como el nivel de digitalización. Con ello se podrá crear grupos homogéneos y personalizados para una gestión más efectiva. Por último, se implementará un modelo de aprendizaje automático para predecir el comportamiento de los nuevos segmentos de clientes y anticipar tasas de recupero.

¹ Nota: se reserva el nombre de la organización bajo estudio por razones de confidencialidad

La importancia de este trabajo radica en su potencial para rediseñar la estrategia de recuperación de deudas en la organización bancaria. La inclusión de la variable de digitalización en la segmentación podría llevar a una mayor eficiencia operativa, en la medida que le permitirá al banco abordar a los clientes morosos con estrategias más personalizadas y efectivas. Además, el uso de modelos de aprendizaje automático para la predicción brindará previsibilidad y permitirá la toma de decisiones informadas.

En resumen, este proyecto busca capitalizar el nivel de digitalización de los clientes para la gestión bancaria, aprovechando la riqueza de los datos disponibles para mejorar la recuperación de deudas. La combinación de la segmentación tradicional con el nivel de digitalización y la aplicación de modelos de aprendizaje automático ofrece el potencial de transformar la manera en que se enfrenta la morosidad, beneficiando tanto a la entidad bancaria como a sus clientes.

Palabras clave

Digitalización bancaria, Gestión de deuda, Segmentación de clientes morosos, *Machine Learning*.

Índice

Resumen	1
Introducción.....	4
1. La digitalización y la gestión eficiente de datos en organizaciones bancarias.....	7
1.1 Análisis de la gestión de datos en el sistema financiero.....	8
1.2 Análisis de datos en la gestión de recuperación de morosidad.....	9
1.3 La clusterización de clientes y su impacto en la gestión de la morosidad.....	11
2. Optimización en la clusterización de clientes para el desarrollo de estrategias de recupero eficientes.....	12
2.1 Extracción y preprocesamiento de datos.	13
2.2 Aplicación del algoritmo <i>k-means</i> para el armado de clústering.....	19
2.3 Análisis de la nueva clusterización.....	20
3. Métodos predictivos para la mayor previsibilidad de comportamiento de clientes morosos.	22
3.1 Aplicación de modelo de aprendizaje automático para la predicción de comportamiento de los nuevos clústeres digitales.....	23
3.2 Análisis y evaluación de resultados.....	27
3.3 La predicción de comportamiento y sus beneficios en el recupero de créditos	31
Conclusión.....	32
Referencias bibliográficas	35

Introducción

La nueva era del Big Data, caracterizada por la disponibilidad de enormes volúmenes de datos, trajo consigo numerosas oportunidades y desafíos para las organizaciones bancarias. Dichas entidades utilizan sus fuentes de datos para, entre otras cosas, mejorar su capacidad de tomar decisiones basadas en evidencia y maximizar la eficiencia de sus esfuerzos de recupero de clientes morosos. Gracias al análisis de datos, los bancos pueden identificar patrones de comportamiento, segmentar a los clientes según su riesgo y personalizar estrategias de recupero más efectivas.

En el contexto planteado, el objetivo de este proyecto es optimizar la segmentación de clientes bancarios morosos de una entidad financiera argentina². Para ello, se busca incorporar el nivel de digitalización de los clientes en el proceso de clusterización y luego predecir el comportamiento que se obtiene de cada clúster. Se considera que el valor agregado generado contribuirá a la redefinición de la estrategia de recupero de clientes morosos en la organización.

En la actualidad, los clientes de las entidades bancarias ingresan en mora cuando no logran pagar la deuda contraída dentro de la fecha de vencimiento establecida. Es así, que los bancos deben gestionar la morosidad de estos clientes a fines de recuperar la deuda que no fue abonada en tiempo y forma. Para ello, los bancos desarrollan diversas estrategias a fines de abordar a los clientes morosos y recuperarlos de la manera más eficiente, es decir, incurriendo en el menor costo posible.

Dentro de la entidad bancaria seleccionada para el presente trabajo, la gestión impartida sobre los morosos se recuesta sobre una segmentación basada en *score*. Este se obtiene a partir de un modelo que tiene en cuenta variables como: ingreso inferido, zona geográfica, grupo etario y la deuda total de cada cliente. Es así como los clientes se segmentan con 3 tipos de riesgos: alto, medio y bajo y se le da a cada aglomeración un tratamiento específico, entendiendo que aquellos clientes de riesgo alto serán los más difíciles de recuperar y los de riesgo bajo los más fáciles. A partir de estos riesgos es que se aborda a los clientes y se decide, entre otras cosas, el día óptimo de asignación a centros de atención telefónica, el día de envío de notificaciones con recordatorios de impago y los honorarios del *Call Center*.

² Nota: se reserva el nombre de la organización bajo estudio por razones de confidencialidad

Ahora bien, esta forma de agrupamiento no tiene en cuenta información que puede resultar útil respecto al nivel de digitalización que tienen sus clientes. La digitalización está comprendida por la cantidad de inicios de sesión o usos de los distintos canales digitales del banco, tales como *App* del teléfono, *Homebanking*, cajeros automáticos y aperturas de *Mail*. Dicha información resultará de vital importancia a la hora de segmentar a los clientes dado que se podrán consolidar perfiles personalizados para así gestionarlos de una manera asertiva, es decir, incurriendo en menores gastos como por ejemplo evitando asignar al *Call Center* a quienes no hiciera falta. De esta manera, el interrogante que surge para llevar adelante este trabajo es: ¿Cómo influye el nivel de digitalización de los clientes de la organización para lograr una gestión más eficiente de la cartera de los morosos?

El presente trabajo tiene, como objetivo, incluir el nivel de digitalización de los clientes para realizar una nueva segmentación de cartera de la organización bancaria. A partir de ello implementar modelos de aprendizaje automático para alcanzar una gestión de la mora más eficiente en el banco. Para lograr desarrollarlo se plantean tres objetivos específicos. Con el primero se busca identificar patrones en el nivel de digitalización de los clientes en el marco de una gestión eficiente de datos sobre la cartera de la organización bancaria. En un segundo objetivo específico, se implementará un modelo de *Clustering* para construir nuevas segmentaciones de clientes por periodo considerando un nuevo atributo que es el nivel de digitalización. Por último, como tercer objetivo, a partir de los resultados obtenidos en los dos anteriores, se implementará un modelo de aprendizaje automático predictivo para lograr predecir el comportamiento de los nuevos clústeres de clientes. Esto con el fin de obtener previsibilidad de la mora y de este modo rediseñar la estrategia de recupero de clientes de la organización bancaria.

Para abordar cada objetivo, el trabajo se divide en tres capítulos. Ante la creciente importancia de comprender los procesos y metodologías de gestión de datos en el sector bancario, en el primer capítulo se llevará a cabo un análisis teórico de diversos enfoques bibliográficos para explorar las distintas perspectivas que definen una gestión eficiente de datos en organizaciones bancarias. De este modo se obtendrá una visión integral y fundamentada sobre cómo abordar de manera efectiva el manejo de

datos en el ámbito bancario, considerando las mejores prácticas, estrategias y experiencias compartidas en la literatura especializada.

En el segundo capítulo, en una primera instancia se realizará la recolección de información necesaria para el análisis. Para ello se recurrirá a fuentes de datos de la organización bancaria y a través de lenguaje SQL se realizarán las extracciones información relevante. Se incluirán datos asociados a la deuda del cliente, el score, nuevas variables de digitalización, entre otras, para el periodo comprendido entre el año 2022 y parte de 2023. Una vez obtenida esta base de datos enriquecida, se construirá un modelo de *Clustering* a través del lenguaje de programación *Python*³ cuyas variables de entrada serán las mencionadas anteriormente. Se logrará, entonces, conformar nuevos segmentos más personalizados con características en común.

A partir de la nueva segmentación creada, en el tercer capítulo se entrenará un modelo de aprendizaje supervisado utilizando información de periodos anteriores y se procederá a predecir tasas de recupero esperadas de cada clúster. Esto repercutirá positivamente en la gestión de los clientes, ya que permitirá determinar, entre otras cosas, nuevos targets de recupero, ofrecerá mayores alternativas a la hora de abordarlos y dará mayor previsibilidad en cuanto a la mora.

³ Acerca de Python: <https://www.python.org/>

1. La digitalización y la gestión eficiente de datos en organizaciones bancarias.

Los autores Ming Syan Chen, Jiawei Han y Philip S. Yu mencionan en su artículo “*Data Mining: An Overview from a Database Perspective*” que el proceso de extracción de conocimiento valioso, regularidades e información relevante de grandes conjuntos de datos, tiene un papel fundamental en la generación y verificación de información. Estos datos se exploran desde diversas perspectivas, convirtiendo a las bases de datos extensas en fuentes confiables de conocimiento. También sostienen que investigadores y empresas reconocen la importancia de extraer información y conocimiento de bases de datos extensas para su aplicación en la toma de decisiones, gestión de información y otros campos dentro del negocio.

En un entorno empresarial multicanal, las compañías están siendo inundadas con datos, lo que abre la posibilidad de, a través de la analítica, generar información para ~~de~~ comprender, gestionar y explotar estratégicamente la dinámica compleja del comportamiento del cliente. Esta práctica se ha vuelto ubicua en la vida diaria. Ejemplos cotidianos demuestran cómo se emplea en diversos campos, desde recibir catálogos personalizados basados en el análisis de comportamientos de compra previos, hasta la evaluación de patrones de comportamiento financiero para predecir la probabilidad de incumplimiento de un préstamo. Desde las redes sociales hasta las transacciones con tarjetas de crédito también son analizadas para predecir comportamientos futuros y adaptar estrategias de negocios en consecuencia (Baesens, 2014).

El objetivo de este apartado es explorar en detalle la relevancia de la digitalización en la industria bancaria y su impacto en la gestión de datos. El avance del *Big Data* ha revolucionado la forma en que los bancos recopilan, procesan y analizan información sobre sus clientes. Esto les ha permitido comprender sus comportamientos y necesidades financieras de manera más precisa. En este contexto, se analizará cómo la aplicación de técnicas de explotación de datos se ha convertido en un pilar fundamental para el sector bancario, en su búsqueda de revelar información valiosa y tomar decisiones estratégicas informadas.

1.1 Análisis de la gestión de datos en el sistema financiero

La digitalización ha tenido un impacto significativo en la industria bancaria. En la actualidad, los bancos pueden recopilar una amplia variedad de información sobre sus clientes, como transacciones y actividades en línea. Esta información es sumamente valiosa, ya que les permite comprender mejor a sus clientes y sus necesidades financieras. Mediante el uso de tecnologías de *Big Data*, los bancos pueden procesar y analizar estos datos para identificar patrones de comportamiento, preferencias y tendencias. Esto les brinda la capacidad de ofrecer servicios más personalizados y adaptados a las necesidades individuales de cada cliente.

El concepto de *Big Data* ha generado una revolución en el mundo empresarial, como bien afirma Schmarzo (2013). Aquellas empresas que adoptan estas tecnologías como una estrategia de transformación pasan de un enfoque retrospectivo a uno prospectivo y predictivo en sus operaciones. Esta tendencia en el avance tecnológico ha abierto nuevas oportunidades para comprender y tomar decisiones de manera más eficiente y efectiva, tal como lo señala Barranco Fragoso (2012).

Descubrir la información oculta en los grandes volúmenes de datos mediante técnicas de Minería de Datos se ha convertido en una tendencia emergente y un objetivo último para una amplia gama de estudios. La banca ha sido un campo de implementación popular de estas para investigadores con habilidades en *Data Mining* a lo largo de las décadas de la revolución de la ciencia de la información. Los bancos han reconocido que el conocimiento, en lugar de los recursos financieros, es el activo más importante. Además, el desarrollo y la popularización de la banca electrónica y la banca móvil contribuyen al crecimiento exponencial de la información bancaria en tiempo real. Estos desarrollos continuos y la creciente disponibilidad de *Big Data* hacen que dominar las herramientas relevantes de análisis en este campo sea una de las tareas más cruciales para el sector bancario (Hassani, Huang y Silva, 2018)

Denis Ostapchenya en su artículo “*The Role of Big Data in Banking: How do Modern Banks Use Big Data?*” confirma que el *Big Data* está desempeñando un papel crucial en el sector bancario, permitiendo que incluso las empresas no bancarias ingresen al ámbito financiero debido a la disponibilidad de datos necesarios y que el *Big Data* en *FinTech* también está transformando la forma en que los propios bancos

operan. Menciona que el cambio en la actitud de los clientes hacia los bancos es evidente, ya que ahora buscan respuestas en línea en lugar de visitar sucursales físicas.

En este artículo, el autor sostiene que las sucursales bancarias están evolucionando para centrarse en otras tareas importantes, mientras que los clientes utilizan aplicaciones móviles y acceden a sus cuentas en línea. La predisposición actual de las personas a compartir información personal también está aumentando el valor del *Big Data*. Las reseñas, la geolocalización y las redes sociales generan una gran cantidad de información, lo que aumenta el rol del *Big Data*. Es así como los bancos están utilizando el *Big Data* para segmentar a sus clientes según sus ingresos y gastos, permitiendo servicios personalizados. Además, la evaluación de riesgos y la prevención de fraudes se optimizan mediante la identificación de patrones de comportamiento financiero. Las opiniones públicas sobre los bancos también se analizan a través del *Big Data*, permitiendo una mejor gestión de comentarios y aumentando la lealtad del cliente.

Bajo este contexto, el análisis de datos y la comprensión del comportamiento de los clientes ofrecen numerosas oportunidades. Los bancos tienen acceso a datos precisos y actualizados que les permiten identificar patrones de comportamiento, evaluar el riesgo crediticio de los clientes y tomar decisiones informadas para gestionar eficientemente, entre otras cosas, la morosidad. Al analizar variables relevantes como historial crediticio, ingresos estimados y otros indicadores financieros, pueden diseñar estrategias personalizadas y asertivas para la recuperación de deudas.

1.2 Análisis de datos en la gestión de recuperación de morosidad

En las empresas que proporcionan servicios de préstamos financieros, es crucial contar con suficiente liquidez para cumplir con todas sus obligaciones. Esto les permite mantener un flujo constante de efectivo que les permita seguir ofreciendo el servicio. Además, estas empresas también se esfuerzan por reducir el riesgo de impago, ya que esto es una característica inherente a la naturaleza de su actividad (Borrero Tigreros y Bedoya Leiva, 2020)

La recolección histórica de datos en el sector bancario se centra en mejorar las estrategias de recupero de clientes que han incurrido en morosidad. Mediante un posterior análisis de estos se busca obtener una predicción más asertiva del

comportamiento del cliente moroso. En la organización bajo estudio, la cartera en mora representa aproximadamente un 4% mensual de las financiaciones totales. Los clientes pueden incurrir en mora con diferentes productos de riesgo y los más usuales son tarjetas de crédito y préstamos.

A partir del primer día en que un cliente entra en morosidad, comienzan a generarse intereses por la deuda impaga. Esto impacta negativamente en la relación banco – cliente. Para gestionarlos, la organización bancaria utiliza diversas estrategias con el objetivo de recuperarlos. El banco categoriza a sus clientes en lo que se conoce como "cajones de mora", que se establecen según la cantidad de días de retraso acumulados por el cliente desde la fecha de inicio de la deuda.

Los "cajones de mora" coinciden con las clasificaciones asignadas por el Banco Central de la República Argentina (BCRA) a los clientes morosos a medida que aumentan sus días de retraso. La clasificación A representa una situación normal, sin retrasos o con un retraso de hasta 30 días. La clasificación B incluye clientes que requieren atención especial, con retrasos de entre 31 y 90 días. La clasificación C se refiere a clientes con problemas, con retrasos de entre 91 y 180 días. La clasificación D se refiere a clientes con un alto riesgo de insolvencia, con retrasos de entre 181 y 365 días. La clasificación E se asigna a clientes considerados irrecuperables, con retrasos superiores a los 365 días. La clasificación F implica una situación irrecuperable según las disposiciones técnicas del BCRA⁴. Para el presente trabajo se toman clientes de la organización que se encuentran en la categoría A.

Con el fin de lograr una gestión más eficiente de la morosidad, la organización bancaria estudiada realiza una segmentación inicial de los clientes morosos basada en dos variables: deuda y *score*. Este *score* es generado por un modelo de aprendizaje automático que tiene en cuenta variables como ingreso inferido, grupo etario y zona geográfica. Esta segmentación inicial divide a los clientes morosos en tres grupos principales: alto riesgo, riesgo medio y bajo riesgo. Cada grupo se aborda con una estrategia específica y los esfuerzos destinados a recuperarlos están estrechamente ligados al riesgo asignado. Es así, que los clientes clasificados como riesgo alto requerirán mayores esfuerzos y costos para ser recuperados, mientras que los de riesgo bajo serán más fáciles de recuperar y, por lo tanto, menos costosos.

⁴ La información del BCRA se encuentra disponible en la siguiente web: <https://www.bcra.gob.ar>

Para realizar una nueva clusterización de clientes que tenga en cuenta no solo la deuda y el score, sino también su nivel de digitalización, se utilizará Teradata-SQL⁵ para acceder a diversas fuentes de datos de la organización bancaria. Estas fuentes de datos contienen información relevante sobre el uso de los 3 canales digitales que ofrece el banco: *Online Banking*, *App* y correos electrónicos. La información sobre el estado de morosidad de los clientes se encuentra almacenada en tablas en un *Data Warehouse*. Con la inclusión de esta información, se realizarán nuevas agrupaciones de clientes a través de la implementación de un modelo de *Clustering* entrenado con el algoritmo *k-means*. El período utilizado para la extracción de información será el año 2022. Además, se anonimizará la información de los clientes, ya que estas fuentes de datos contienen información sensible.

1.3 La clusterización de clientes y su impacto en la gestión de la morosidad

En el contexto mundial de digitalización, autores como Yuping, Jilková, Guanyu y Weisl (2020), invitan a pensar en nuevos métodos de segmentación o clusterización de clientes. Plantean que la segmentación juega un papel crucial en el desarrollo de estrategias comerciales rentables y que los métodos comunes de segmentación incluyen variables demográficas, geográficas, psicográficas y de comportamiento. Recomiendan a nuevos investigadores que han explorado la segmentación basada en patrones de transacción, técnicas de minería de datos y algoritmos de aprendizaje automático.

También resulta interesante el enfoque de segmentación planteado por Mousaeirad (2020). En su artículo “*Intelligent Vector-based Customer Segmentation in the Banking Industry*” aborda la importancia de la segmentación inteligente de clientes en la industria bancaria y explica cómo el uso de técnicas de aprendizaje automático y minería de datos puede mejorar su comprensión y la toma de decisiones estratégicas en los bancos. El trabajo presenta un aporte relevante en el campo de la calificación crediticia al abordar la diferenciación de dos tipos de clientes morosos: aquellos que no pueden pagar debido a problemas de flujo de efectivo y aquellos que no quieren pagar debido a una falta de voluntad. Esta diferenciación proporciona una comprensión más profunda del comportamiento de los morosos y permite una clasificación más precisa de los prestatarios.

⁵ Acerca de Terada-SQL: <https://www.teradata.com/>

En esta línea, autores como Bravo, Thomas y Weber (2015), proponen incorporar variables dinámicas al análisis de datos para explicar el comportamiento del cliente deudor a lo largo del tiempo en pos de mejorar el modelo de *Scoring* crediticio tradicional. De la misma manera, Yap, Ong y Husain (2011), recomiendan la utilización de técnicas de minería de datos para mejorar la evaluación de la solvencia crediticia a través de estos modelos de *Scoring*.

A su vez, Verdenhofs y Tambovceva (2019), explican el impacto que los avances en la tecnología y la gran oferta disponible de datos generan en la forma en que las empresas segmentan a sus clientes. En su artículo, afirman que el enfoque tradicional de segmentación basado sólo en características demográficas y socioeconómicas ya no es suficiente para comprender plenamente las necesidades y comportamientos de los clientes. A raíz de esto, sugieren que el uso de datos masivos y técnicas de análisis avanzadas ofrecen nuevas oportunidades para segmentar a los clientes de manera más precisa y efectiva.

En todos los casos citados, se puede destacar la relevancia de utilizar toda la información disponible en cada organización. El uso de los datos históricos almacenados por cada institución financiera es absolutamente relevante para ser más eficientes en la gestión de recupero. A partir de ello, la institución financiera podrá generar perfiles de clientes cada vez más reales y así podrá predecir su comportamiento de una manera cada vez más asertiva. Gracias a esto logrará ofrecer al cliente opciones o alternativas de pago más personalizadas que permitirán a la organización bancaria lograr mejores resultados de gestión.

2. Optimización en la clusterización de clientes para el desarrollo de estrategias de recupero eficientes

Como se mencionó en el capítulo previo, la segmentación de clientes de la entidad bancaria seleccionada se basa únicamente en las variables deuda y score. En este apartado, a partir de obtener la información correspondiente al uso de los medios digitales de los clientes se incluye como una nueva variable que representa el nivel de digitalización de cada uno. A partir de ello se construye una nueva segmentación. Los cortes temporales de los datos serán mensuales debido a que es la periodicidad con la que se encuentran disponibles.

De esta manera, para cada mes del año 2022 y parte del 2023, se contará con un conjunto de clientes anonimizados que ingresaron en mora y se utilizarán las variables deuda, score y cantidad de usos de canales digitales por mes. Partiendo de esta información y aplicando el algoritmo de aprendizaje no supervisado *k-means*, se entrenará un modelo para realizar una clusterización de estos clientes. Se configurarán 3 grupos con diferentes niveles de digitalización.

Esta nueva segmentación proporcionará agrupaciones de clientes con características aún más similares que las segmentaciones utilizadas actualmente por la entidad bancaria. Esto permitirá abordarlos con estrategias de recupero más personalizadas en base a esta característica. Además, se espera que los grupos de clientes más digitales obtengan tasas de recupero más altas, mientras que aquellos con menor grado de digitalización tengan tasas más bajas. De este modo, la organización bancaria podrá obtener una mayor previsibilidad, lo que se traducirá en esfuerzos de gestión más efectivos y menos costosos.

2.1 Extracción y preprocesamiento de datos.

Los datos con los cuales se trabajará fueron obtenidos del *Data Warehouse* de la entidad bancaria bajo análisis. Se emplearon las bases de datos de la organización, las cuales albergan diferentes datos sobre los clientes, dentro del entorno productivo proporcionado por la herramienta Teradata. A través del uso del lenguaje SQL, se llevaron a cabo las consultas necesarias para extraer la información relacionada con las deudas que están bajo gestión de la organización.

Las variables relevantes obtenidas para el análisis por cada cliente que ingresó a mora durante el periodo estudiado son:

Tabla 1. Descripción de variables

Variable	Descripción
<i>Periodo</i>	Mes y año de ingreso a mora del cliente.
<i>Party_id</i>	Es un identificador único para cada persona.
<i>Dda_vda</i>	Deuda vencida del cliente
<i>Dda_Tot_Id</i>	Deuda total del cliente (incluye deuda vencida y deuda a vencer).
<i>Score</i>	Valor numérico arrojado por un modelo del banco que incluye variables como ingreso inferido, deuda, actividad, entre otros.
<i>Flag_recuperado</i>	Variable binaria que muestra 1 si el cliente fue recuperado y 0 si el cliente no fue recuperado.
<i>APP</i>	Cantidad de inicios de sesión a través de la APP por mes de cada cliente.
<i>HB</i>	Cantidad de inicios de sesión a través de HomeBanking por mes de cada cliente.
<i>Banelco</i>	Cantidad de logueos en cajeros automáticos por mes de cada cliente.
<i>Cant_Log_Tasi</i>	Cantidad de logueos en Terminales de autoservicio por mes de cada cliente.
<i>Cant Envio Mail Ok Tot</i>	Cantidad de envíos de mail por mes de cada cliente.
<i>Cant Lect Mail Tot</i>	Cantidad de aperturas de mail por mes de cada cliente.
<i>Activity_code</i>	Es un código que representa la actividad que ejerce cada cliente.
<i>Suc_Madre</i>	Representa la sucursal madre a la que está relacionada el cliente.
<i>Profession_Code</i>	Es un código que indica la profesión de cada cliente.
<i>Edad</i>	Representa la edad de cada cliente.
<i>Ing_Inferido</i>	Representa el ingreso inferido de cada cliente.

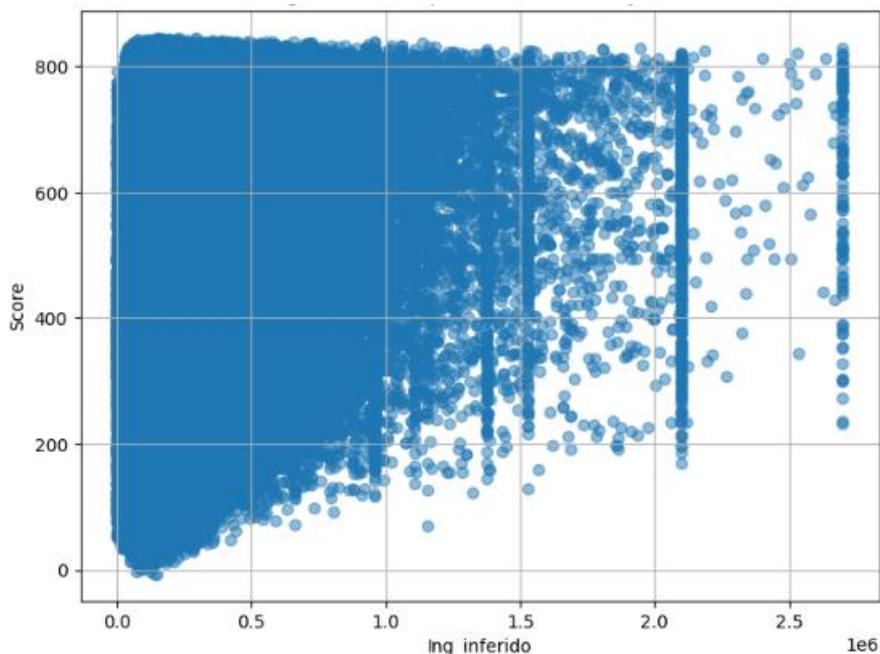
Fuente: Elaboración propia con Excel

Para establecer una base sólida en la comprensión de los datos contenidos en la Tabla 1, en primer lugar, se realiza un análisis exploratorio y descriptivo. La dimensión del conjunto de datos se encuentra conformada por 5.410.198 filas y 17 columnas. Estas últimas son cada una de las variables de la tabla 1, mientras que las filas representan a cada cliente. Se observa que casi el 83% de los clientes corresponden a clientes recuperados ($\text{Flag_recuperado} = 1$).

El análisis descriptivo, permite obtener una visión detallada de las características esenciales del conjunto de datos tales como las tendencias centrales, la dispersión y la presencia de valores atípicos. Este proceso sintetiza la información en estadísticas clave y alerta sobre cualquier anomalía potencial que pueda afectar a los análisis futuros.

Por otra parte, con el análisis exploratorio se obtuvo una comprensión más profunda y significativa de los datos al descubrir patrones y relaciones subyacentes. Se logró explorar visualizaciones más complejas y gráficos que resaltaron conexiones entre variables, permitiendo así formular hipótesis iniciales sobre las interacciones presentes.

Figura 1: Diagrama de dispersión entre Score e Ingreso inferido



Fuente: elaboración propia con Python

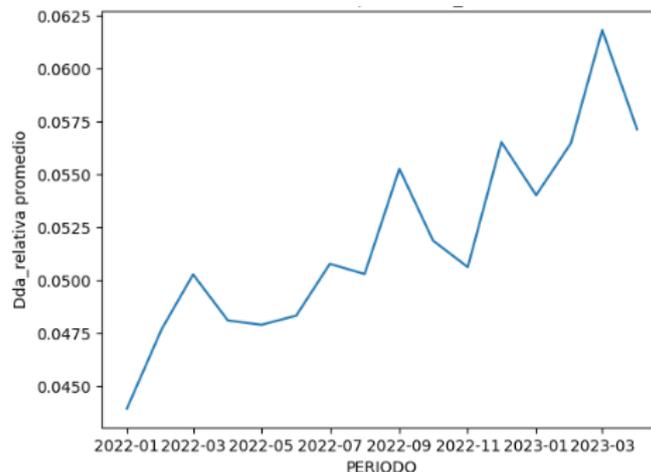
A partir de la figura 1, se observa que los scores más altos poseen ingresos inferidos más elevados. Además, con el análisis exploratorio permitió identificar datos faltantes, lo cual resulta crucial para tomar luego decisiones informadas respecto a cómo abordarlos y garantizar la calidad de nuestros resultados futuros. Gracias a esta etapa se

logró adquirir una visión completa de la base de datos inicial, comprendiendo sus características para poder abordar luego adecuadamente la etapa de preprocesamiento y de modelado.

A continuación de estos primeros pasos se realizó un preprocesamiento de los datos con *Python* a través de diferentes funciones para asegurar la calidad de los datos antes de avanzar con el modelado. En primer lugar, se detectaron e imputaron valores nulos. Para los casos donde se detectó valores “NaN” en las variables de inicios de sesión y cantidad de logueos se las reemplazó con 0 para indicar la inactividad del cliente. Tanto los valores de la columna ‘Dda_vda’ como los de la columna ‘Dda_Tot_Id’ se convirtieron de formato numérico con comas a formato de punto flotante, utilizando la función *replace(',', '.', regex=True)* y luego se convirtieron al tipo de dato *float* mediante *astype(float)*. Se convirtieron los valores en la columna ‘PERIODO’ del DataFrame ‘df’ a formato de fecha y hora, utilizando el formato ‘AñoMes’ (*%Y%m*). Se imputaron los valores nulos en la columna ‘EDAD’, “SCORE” e “ING_INFERIDO” con la media calculada utilizando la función *fillna()* con el argumento *inplace=True* para modificar el *DataFrame* original y se convierten los valores a tipo de dato entero mediante *astype(int)*.

Además, se creó la variable ‘Dda_relativa’ para medir la proporción que representa la deuda vencida sobre la deuda total. Esta proporción de deuda se calculó dividiendo los valores de la columna ‘Dda_Vda’ entre los valores de la columna ‘Dda_Tot_Id’. Se gráfica obteniéndose la Figura 1 donde se observa un crecimiento sostenido de la deuda relativa de los clientes.

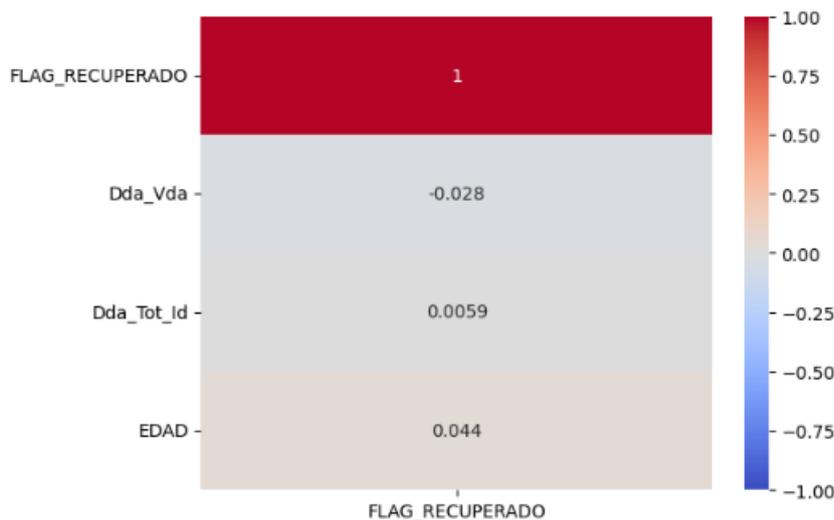
Figura 2. Evolutivo de deuda relativa



Fuente: Elaboración propia con Python

Se crea un mapa de calor para visualizar las correlaciones entre la variable 'Flag_Recuperado' y las variables 'Dda_Vda', 'Dda_Tot_Id' y 'Edad'. Esto permite identificar posibles relaciones lineales entre ellas. Los valores de correlación en la Figura 2 se representan con colores, donde los tonos más claros indican correlaciones más débiles, mientras que los tonos más oscuros indican correlaciones más fuertes. Como resultado se obtuvo que no hay relaciones perfectamente lineales.

Figura 3. Mapa de calor Flag_recuperado

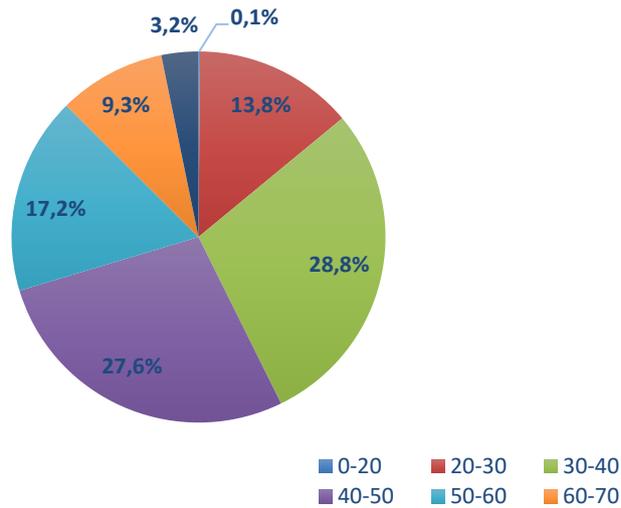


Fuente: Elaboración propia con Python

También, se realizaron transformaciones en algunas variables a fines de simplificar los datos. La variable 'Cant_mail_envios_ok' y 'Cant_mail_lect_ok' se fusionaron a través de un cociente a fines de relativizar la apertura de mail respecto a envíos realizados formando una nueva variable 'Ratio_apertura_mail'. Además, se transformaron a enteros las variables 'Cant_Envio_Mail_Ok_Tot', 'Cant_Lect_Mail_Tot', 'Cant_Log_Tasi', 'Activity_code', 'Suc_Madre', 'Profession_Code' y 'APP' y se eliminaron los caracteres innecesarios.

Por otra parte, se configuró un rango de edad para verificar la distribución de edades en el universo total y se corrobora que los grupos más relevantes corresponden al rango de entre 20 y 60 años como se puede observar en la Figura 4 a continuación.

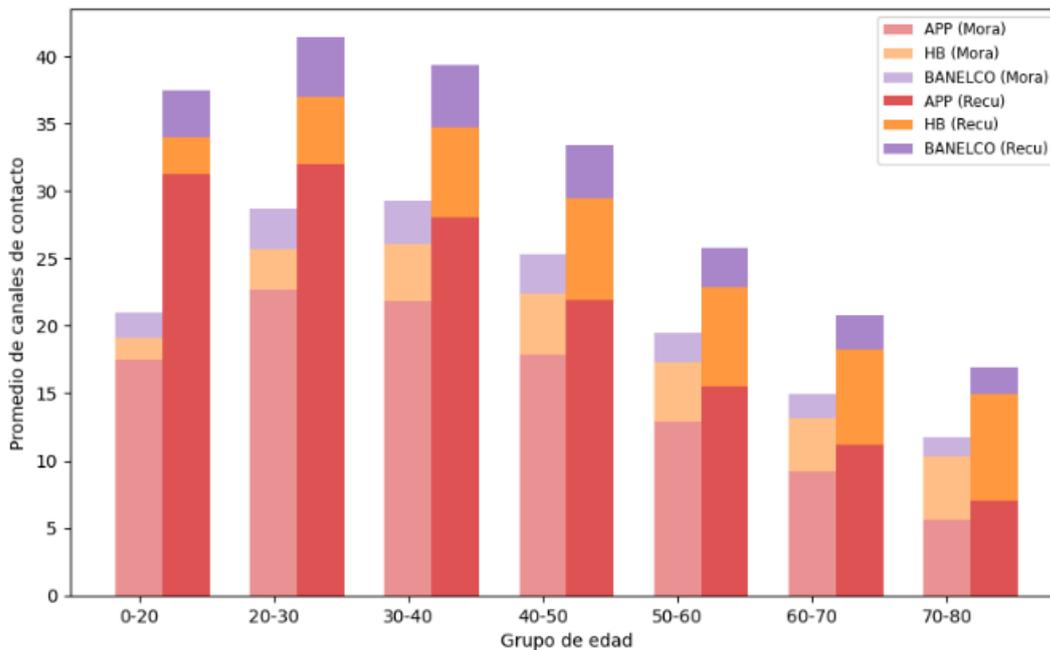
Figura 4. Distribución por rango etario como porcentaje del total



Fuente: Elaboración propia con Excel

Se analizaron por separado los universos de clientes recuperados y los no recuperados agrupando por rango etario a fines de analizar si existen patrones y relaciones subyacentes respecto a la digitalización de estos. Se contabiliza la cantidad de inicios de sesión de las variables APP, HB y BANELCO para comparar niveles de digitalización de cada uno y se visualiza la información.

Figura 5. Digitalización promedio según rango etario, canal y recupero

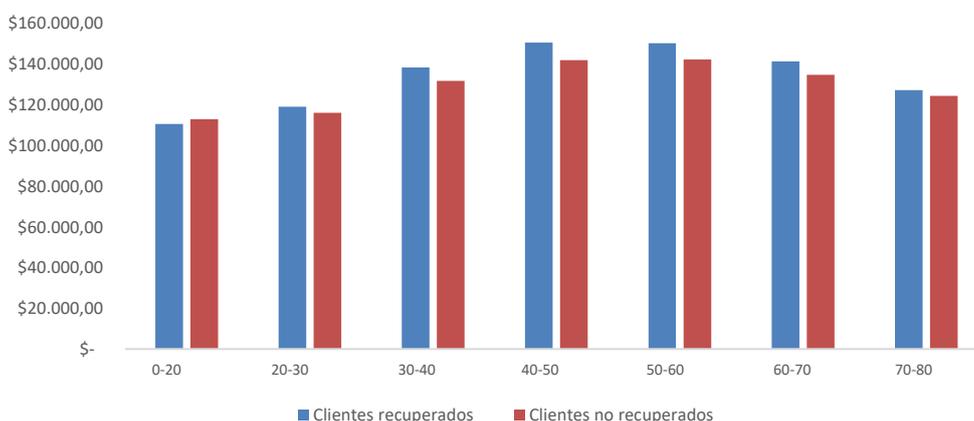


Fuente: Elaboración propia con Python

A partir de la figura 5, se obtienen dos observaciones. Por un lado, que los grupos más jóvenes utilizan los canales digitales en mayor medida que los grupos más adultos. Por el otro, que los clientes recuperados presentan mayor utilización de los canales digitales que los clientes que no se recuperan. Estas observaciones son relevantes ya que sugieren que el banco puede mejorar sus estrategias de recupero al dirigirse de manera más efectiva a los grupos más jóvenes y capitalizar la disposición de los clientes recuperados para utilizar canales digitales, lo que podría aumentar la eficacia general en el recupero de clientes morosos.

En la Figura 6 a continuación se puede observar una comparativa de la media de ingresos inferidos agrupado por rango etario entre clientes recuperados y no recuperados.

Figura 6. Ingresos inferidos según rango etario y recupero



Fuente: Elaboración propia con Excel

De este modo se verifica que los grupos de entre 40-50 y 50-60 son los que presentan la media más alta de ingresos inferidos. Además, se observa que los clientes recuperados tienen un ingreso medio inferido más alto que aquellos no recuperados.

En conclusión, se seleccionaron variables y a través de un análisis descriptivo, se exploraron tendencias y características del conjunto de datos, identificando valores faltantes y anomalías. Se realizó un preprocesamiento que incluyó la imputación de valores nulos y transformaciones de datos para una mejor calidad. Además, se analizó la relación entre variables, como la proporción de deuda vencida y total. Se observó un aumento constante en la deuda relativa. Los datos se simplificaron, y se descubrieron

patrones como la mayor adopción digital en grupos más jóvenes y en clientes recuperados.

2.2 Aplicación del algoritmo *k-means* para el armado de clústering.

K-Means es un algoritmo de aprendizaje no supervisado que tiene la capacidad de clasificar datos en grupos o "*clústers*" similares. Su principal objetivo es encontrar conjuntos de datos que compartan similitudes entre sí, formando agrupaciones distintas y bien definidas. Su funcionamiento se desarrolla en varias etapas.

El proceso de *k-means* implica seleccionar algunos puntos como los puntos focales iniciales del clúster (generalmente, se eligen los primeros K puntos de muestra de ingresos); luego, se agrupan los puntos de muestra restantes en sus puntos focales de acuerdo con el criterio de distancia mínima, obteniendo así una clasificación inicial. Si la clasificación no es razonable, se modifica, y se repite iterativamente hasta obtener una clasificación razonable (Youguo Li y Haiyan Wu, 2012).

En la aplicación realizada se lleva a cabo una serie de pasos para preparar los datos y realizar un análisis de *Clustering* utilizando el algoritmo *K-Means*. Se comenzó por crear un nuevo *DataFrame* denominado "df_digital" a partir de una selección de columnas específicas del *DataFrame* original. Este nuevo *DataFrame* se utilizó como base para el análisis y *Clustering*. Las columnas seleccionadas incluyeron información relacionada con el período, identificación de clientes, *Score*, actividad en aplicaciones móviles (APP), actividad en *Homebanking* (HB), ingreso inferido (Ing_Inferido), interacciones con el cajero automático y terminal de autoservicio "BANELCO" y "CANT_LOG_TASI", ratio de apertura de correos electrónicos (ratio_mail), deuda total (Dda_Tot_Id), edad (EDAD) y la variable objetivo "FLAG_RECUPERADO".

Posteriormente, se convirtieron diversas columnas del *DataFrame* en tipo de dato enteros o *floats* para asegurar su correcta interpretación en el análisis. Esto incluyó las columnas "Hb", "Banelco", "App", "Cant_Log_Tasi", "Cant_Lect_Mail_Tot" y "Cant_Envio_Mail_Ok_Tot". Adicionalmente, se sumaron las columnas "Banelco" y "Cant_Log_Tasi" para obtener una nueva columna llamada "Atm-Tasi". Esta columna representa la suma de las interacciones en cajeros automáticos y terminales de autoservicio.

Finalmente, se realizó una normalización de los datos utilizando la función *StandardScaler* de la biblioteca *sklearn.preprocessing* en *Python*. La normalización ayuda a que las variables tengan una escala similar, evitando que las variables con escalas más grandes dominen el proceso de *Clustering*. Luego, se aplicó el algoritmo *K-Means* al conjunto de datos normalizado. Se eligieron 3 clústeres (grupos) para agrupar los datos. En el siguiente apartado se aborda el análisis del resultado obtenido.

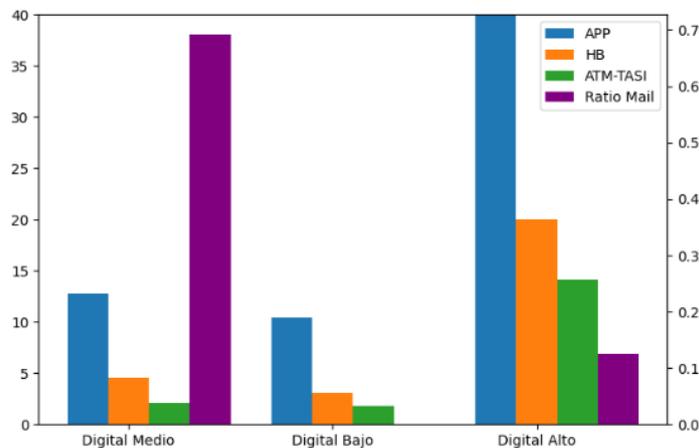
En resumen, se realizó un proceso de preparación de datos y análisis de *clustering* utilizando el algoritmo *K-Means*. Los datos han sido seleccionados, transformados, normalizados y agrupados en clústeres. Esto va a permitir explorar patrones y relaciones dentro de los datos y generar segmentaciones que pueden ser útiles para tomar decisiones informadas en el contexto de la organización bancaria.

2.3 Análisis de la nueva clusterización.

Luego de la clusterización realizada en el apartado anterior, se verifican las medias y medianas de cada grupo obtenido. Para ello se comienza agrupando los datos con la columna 'clúster' que fue generada durante el proceso de *Clustering*. Luego, se calculan las medias de las variables 'APP', 'HB', 'ATM-TASI' y la mediana de 'ratio_mail' para cada grupo de clúster. Los valores calculados se dividen por un millón para tener una escala más manejable y expresar los valores en millones.

Se renombran los clústeres numéricos por etiquetas más descriptivas utilizando el método '*rename*' en el índice del *DataFrame* "*grouped_data*" y se crea un gráfico de barras agrupado utilizando la función '*plot*' en *Python*.

Figura 7. Digitalización por clúster

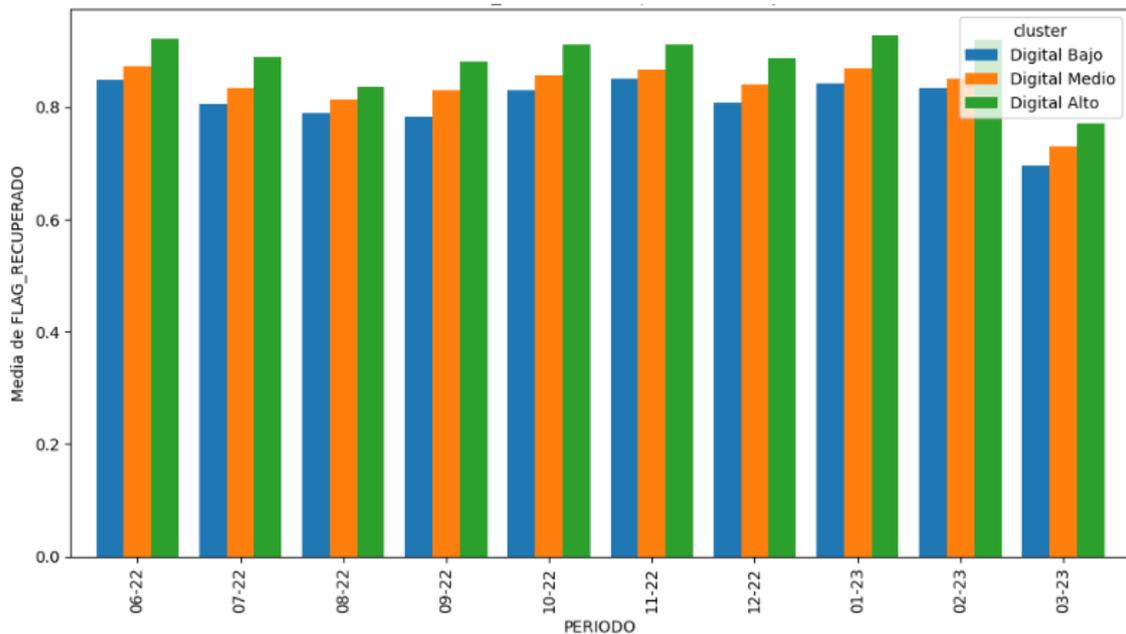


Fuente: Elaboración propia con Python

En la Figura 7, se puede observar un clúster con clientes con altos índices de digitalización donde se destaca el uso de APP, *HomeBanking* y ATM-TASI. También, un clúster de clientes con índices de digitalización media donde se destaca sobre todo el uso de *Mail* y un clúster de clientes con índices de baja digitalización donde se registran bajos niveles de utilización de los canales digitales.

A continuación, se calculan las tasas de recupero de cada clúster obtenido (deuda recuperada / deuda total). En la Figura 8 se visualiza la tasa de recupero de cada clúster en el periodo estudiado. De esta, se obtiene la valiosa apreciación que a mayor digitalización mejor comportamiento de los clústeres en términos de recupero.

Figura 8. Recupero por clúster



Fuente: Python – Elaboración propia

Además, se obtuvo que el *score* más digital contiene 975.935 valores con una media de *score* de 572, el *clúster* con digitalización media contiene 1.120.518 valores con una media de *score* de 484 y el *clúster* menos digital contiene 3.313.745 con una media de *score* de 476. Los valores registrados de *score* se condicen con los índices de recupero analizados dado que se espera que a menor *score* peor comportamiento en términos de recupero. El recupero promedio de cada grupo es de 89% el *clúster* digital alto, 83% el *clúster* digital medio y de 81% el *clúster* digital bajo.

A partir del análisis realizado se puede concluir que se logró clasificar a los clientes de la organización bancaria en grupos con características similares respecto a su

nivel de digitalización. Gracias a esto es posible percibir de manera más certera el comportamiento del cliente pudiendo analizar cada grupo de manera particular.

3. Métodos predictivos para la mayor previsibilidad de comportamiento de clientes morosos.

El análisis predictivo es un área de la minería de datos que consiste en la extracción de información existente en los datos y su utilización para predecir tendencias y patrones de comportamiento (Timón C., 2017). A través de los métodos predictivos, las organizaciones o industrias pueden identificar los patrones dentro de los datos y hacer pronósticos futuros sobre la base de datos y técnicas de análisis existentes (Rustagi y Goel, 2022). Los modelos predictivos son herramientas que establecen una relación entre el rendimiento de una unidad en una muestra y sus atributos conocidos.

Eric Siegel, en su libro *“The Power to predict who will click, buy, lie or die”*, afirma que la predicción desempeña un papel fundamental en una variedad de campos, desde la meteorología hasta la toma de decisiones empresariales. El texto introduce el campo de la analítica predictiva, resaltando su importancia en la toma de decisiones basadas en pronósticos sobre el comportamiento futuro de individuos. Se subraya que las organizaciones pueden beneficiarse enormemente de la analítica predictiva, mejorando la eficiencia, tomando decisiones más informadas. Se enfatiza, también, la importancia de la predicción en diversas áreas y su capacidad para mejorar la toma de decisiones y los resultados en diferentes campos de aplicación, tales como predicción de precios de acciones, riesgos, marketing, entre otros.

En el análisis predictivo se construye un modelo estadístico utilizando datos existentes para predecir datos desconocidos. Se utilizan conjuntos de entrenamiento con atributos y comportamientos conocidos, y conjuntos de prueba con atributos similares, pero comportamientos desconocidos. El análisis predictivo incluye líneas de tendencia y puntuación de influencia, entre otros (Timón, C., 2017).

En el contexto de la gestión de clientes morosos en instituciones financieras, los métodos predictivos desempeñan un papel fundamental para obtener una mayor previsibilidad del comportamiento de estos clientes y tomar decisiones más informadas y eficaces que ayuden a priorizar las acciones de recuperación. Bajo este enfoque se

entrenará un modelo de aprendizaje automático con la información histórica recolectada a fines de predecir de manera más asertiva el comportamiento de los clientes.

3.1 Aplicación de modelo de aprendizaje automático para la predicción de comportamiento de los nuevos clústeres digitales.

Para poder llevar a cabo el entrenamiento de un modelo de aprendizaje supervisado primero se deben reunir y preparar los datos, asegurando que estén limpios y listos para su procesamiento. Luego, se divide el conjunto de datos en entrenamiento y prueba para entrenar el modelo y se elige qué tipo de modelo se adaptará mejor al problema en cuestión. El mismo se entrena con los datos de entrenamiento ajustando sus parámetros internos para, finalmente, evaluar su rendimiento utilizando el conjunto de prueba a través de diversas métricas según el tipo de problema. Dado que el conjunto de datos seleccionados para el presente trabajo presenta una clase mayoritaria (clientes recuperados) el modelo podría sesgarse a la hora de entrenarse y es por esto por lo que se utilizará un algoritmo *Adaboost*. Este resulta ideal para este tipo de datos ya que combina múltiples modelos y les otorga más importancia a los errores cometidos dentro de la clase minoritaria.

AdaBoost, abreviatura de "*Adaptive Boosting*", es un algoritmo de aprendizaje automático utilizado en problemas de clasificación y regresión. Fue propuesto por Yoav Freund y Robert Schapire en 1996 y fue el primer algoritmo de *boosting* práctico y sigue siendo uno de los más ampliamente utilizados y estudiados, con aplicaciones en numerosos campos.

La idea fundamental detrás de *AdaBoost* es combinar múltiples modelos de aprendizaje débiles (también llamados "clasificadores débiles") para construir un modelo más fuerte y preciso. Un clasificador débil es un modelo que tiene un rendimiento ligeramente mejor que el azar, por lo que su tasa de aciertos es un poco mejor que el 50% (en el caso de clasificación binaria). *AdaBoost* se enfoca en mejorar la precisión de estos clasificadores débiles mediante un proceso de ajuste ponderado (Schapire, 1999).

AdaBoost tiene las ventajas de ser rápido en velocidad, simple en operación y fácil de programar. No es necesario ajustar parámetros excepto por el número de iteraciones. Puede combinarse de manera flexible con cualquier método para buscar hipótesis débiles sin necesidad de conocimiento previo de *WeakLearn*. Dado suficientes

datos y un *WeakLearn* con una moderada precisión confiable, puede proporcionar un conjunto de garantías teóricas de aprendizaje. Se enfoca en buscar un método de aprendizaje débil solo mejor que una predicción aleatoria, en lugar de tratar de diseñar un algoritmo preciso en todo el espacio (Tu Chengsheng, Liu Huacheng y Xu Bing, 2017).

El proceso de *AdaBoost* ocurre en varias etapas. En primer lugar, se asignan pesos a cada ejemplo en el conjunto de entrenamiento, reflejando su importancia relativa. Se entrena un modelo débil con estos datos, priorizando ejemplos con pesos más altos para minimizar el error ponderado. A medida que avanza, se calcula el error ponderado del modelo débil según los ejemplos y sus pesos. Los modelos con menor error obtienen coeficientes de importancia más altos. Luego, se actualizan los pesos de los ejemplos según cómo el modelo los clasificó. Los ejemplos mal clasificados tienen pesos más altos para enfocarse en ellos en la próxima iteración. Se combinan los modelos débiles en uno fuerte, donde cada uno contribuye ponderado por su coeficiente. Este proceso se repite varias veces, mejorando la precisión general. Finalmente, el modelo fuerte se usa para predecir nuevas instancias considerando las contribuciones de los modelos débiles ponderados por sus coeficientes (Freund y Schapire, 1996)

Se lleva a cabo entonces una evaluación exploratoria y la construcción del modelo predictivo con el propósito de comprender y predecir el comportamiento de clientes en situación de morosidad. Inicialmente, se organizan los datos en dos componentes fundamentales: características y etiquetas. Mientras que las características representan las variables que contribuirán a la predicción, las etiquetas contienen los resultados que se aspiran a predecir. Posteriormente, los datos se subdividen en conjuntos independientes de entrenamiento y prueba, permitiendo que el modelo se entrene en uno y evalúe su rendimiento en el otro.

Los hiperparámetros clave del modelo son definidos en esta fase, entre ellos, la tasa de aprendizaje (*learning_rate*) y el número de estimadores (*n_estimators*). Estos parámetros influyen en la forma en que se fusionan los modelos individuales. Una vez establecidos, se procede a generar una instancia del clasificador *AdaBoost*, incorporando los hiperparámetros ya definidos. Posteriormente, el modelo es entrenado utilizando los datos de entrenamiento.

Una vez finalizado el proceso de entrenamiento, el modelo es empleado para llevar a cabo predicciones en los datos de prueba. Luego, se evalúa el rendimiento del modelo a través de la métrica *Precision*, la cual refleja la proporción de predicciones correctas respecto al total de predicciones positivas; la métrica *Accuracy* que proporciona la porción de predicciones positivas correcta respecto del total de predicciones realizadas; la métrica *Recall* que revela la proporción de ejemplos positivos correctamente clasificados al total de casos positivos; la métrica *f1-score* la cual es una medida equilibrada entre *Precision* y *Recall*. También se evalúa la métrica *Support* que representa el número de ocurrencias de cada clase en el conjunto de datos de prueba y la métrica *Weighted_avg* que calcula el promedio ponderado de métricas como *precisión*, *Recall* y *F1-score* para cada clase.

Figura 9. Métricas de modelado

Assessments	Formula
Precision (<i>P</i>)	$\frac{TP}{TP+FP}$
Recall (<i>R</i>)	$\frac{TP}{TP+FN}$
F1-score	$2 \times \frac{P \times R}{P+R}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$

Fuente: tomado de <https://sitiobigdata.com/>

Otra alternativa para evaluar el rendimiento del modelo es la métrica AUC (*Area under the ROC Curve*) que representa el área bajo las curvas ROC (*Receiver operating characteristic*). Estas últimas, son una herramienta valiosa para evaluar el rendimiento de clasificadores en problemas de clasificación binaria, ya que muestran cómo varían las tasas de verdaderos positivos y falsos positivos a medida que se ajusta el umbral de decisión. Las curvas ROC son gráficos bidimensionales donde la tasa de verdaderos positivos se representa en el eje Y y la tasa de falsos positivos en el eje X. El punto (0,0) representa la estrategia de nunca emitir una clasificación positiva, mientras que el punto (1,1) representa la emisión incondicional de clasificaciones positivas. Cada punto en la curva ROC representa un umbral de decisión diferente, siendo este el valor que el modelo utiliza para determinar si una instancia se clasifica como positiva o negativa (Fawcett, 2005).

El AUC es ampliamente utilizado en tareas de clasificación binaria y se calcula bajo suposiciones paramétricas, semiparamétricas y no paramétricas (Wu, 2005). A

través de la representación gráfica AUC se refleja cuán capaz es el modelo de distinguir cada clase. Esta medida varía entre 0 y 1.0, donde 0.5 representa el rendimiento de un clasificador que adivina al azar, y un valor superior a 0.5 indica un mejor rendimiento que el azar (Fawcett, 2005).

Luego de verificar las métricas principales, se lleva a cabo un proceso avanzado de búsqueda de hiperparámetros, con el fin de afinar aún más el rendimiento del modelo. En esta etapa, se seleccionan las características consideradas relevantes y se procede a dividir los datos en conjuntos de entrenamiento y prueba. Una subdivisión adicional del conjunto de entrenamiento en subconjuntos de entrenamiento y validación contribuye a un ajuste más preciso de los hiperparámetros.

A continuación, se define una distribución de valores posibles para los hiperparámetros que se desean ajustar, incluyendo el número de estimadores y la tasa de aprendizaje. El enfoque adoptado implica el uso de "*RandomizedSearchCV*", una técnica de búsqueda aleatoria que persigue la identificación de combinaciones óptimas de hiperparámetros para maximizar el rendimiento del modelo. Los autores James Bergstra y Yoshua Bengio en su artículo "*Random Search for Hyper-Parameter Optimization*" comparan las estrategias de búsqueda en cuadrícula (*grid search*) y búsqueda aleatoria (*Random Search*) para la optimización de hiperparámetros. Los resultados empíricos y teóricos demuestran que las pruebas elegidas al azar son más eficientes para la optimización de hiperparámetros que las pruebas en una cuadrícula.

Tras concluir este proceso, se identifican los mejores hiperparámetros y la puntuación más alta obtenida en términos de rendimiento. El resultado de esta optimización fue una ligera mejora en la precisión general. Se observa una mejora en la métrica *Precision* de la clase 0 pero una pequeña caída del *Recall*.

De todas formas, se procede con la utilización de estos hiperparámetros dado que para la organización bajo análisis es valioso mejorar la métrica *Precision* debido a que mide la calidad de las predicciones positivas. Los falsos positivos son aquellos casos en los que el modelo predice que un evento ocurrió (en este contexto, que un cliente será recuperado), pero en realidad no ocurrió. La métrica *Precision* se reduce cuando hay muchos falsos positivos, ya que esto aumenta el denominador de la fórmula y disminuye la proporción de verdaderos positivos en la predicción. Resulta, entonces, crucial tener en cuenta los falsos positivos debido a que son clientes que se predicen

incorrectamente que pagarán. Esto implicaría despriorizar clientes que necesitan ciertos estímulos para realizar su pago.

En resumen, se emplearon diversas técnicas y herramientas en Python para establecer y refinar un modelo de predicción con la finalidad de comprender y anticipar el comportamiento de clientes en situación de morosidad. La optimización de hiperparámetros, la evaluación exhaustiva del rendimiento y la búsqueda de configuraciones óptimas convergen en la búsqueda de la mayor precisión posible en las predicciones generadas por el modelo.

3.2 Análisis y evaluación de resultados.

Una vez aplicado el modelo se obtuvieron las medidas de performance explicadas en el apartado anterior. Los valores de las métricas obtenidas se pueden apreciar en la Tabla 2 a continuación.

Tabla 2. Métricas del modelo

Métricas	Valores
<i>Accuracy</i>	87,00%
<i>Precision</i>	88%
<i>Recall</i>	97%
<i>F1-score</i>	92%
<i>Support</i>	Clase 0: 184.490 Clase 1: 897.550
<i>weighted avg</i>	85%

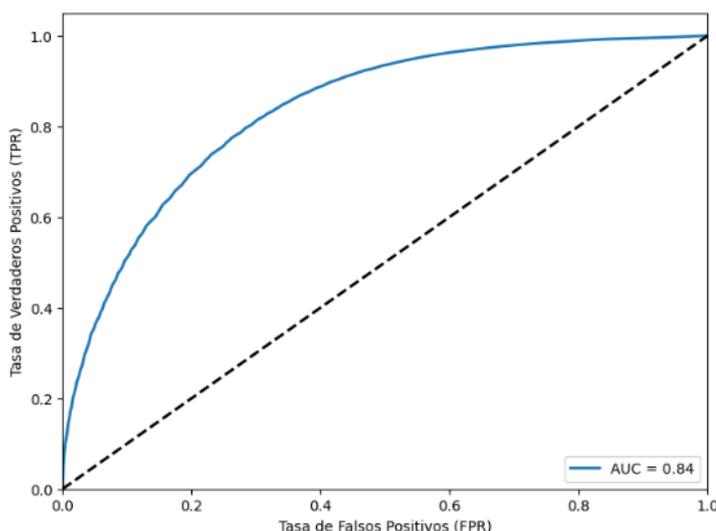
Fuente: Elaboración propia con Excel

El modelo desarrollado alcanzó una *Accuracy* del 87%, lo que significa que pudo clasificar correctamente esa proporción de las instancias en el conjunto de datos evaluado. Por otra parte, el 88% en la métrica *Precision* indica que cuando el modelo predijo que algo pertenecía a esta clase, en la mayoría de los casos fue correcto. La métrica *recall* o tasa de verdaderos positivos obtuvo un 97% para la clase positiva, lo que indica que el modelo fue capaz de capturar casi la totalidad de las instancias reales de esa clase. El modelo logró, además, un *F1-score* del 92% para la clase positiva, lo que indica un buen equilibrio entre la identificación correcta de las instancias positivas y el sorteo de los errores. Finalmente, la métrica *support* refleja que se encontraron 184.490 instancias de la clase 0 y 897.550 instancias de la clase 1 en el conjunto de datos.

Además de estas métricas específicas, es importante destacar que la precisión ponderada (*weighted avg*) fue del 85%, lo que refleja un buen rendimiento general del modelo al considerar todas las clases y tomar en cuenta su distribución en el conjunto de datos. Si bien el modelo ha obtenido resultados alentadores, siempre hay margen para mejoras y refinamientos. Las métricas *precision*, *recall* y *F1-score* para la clase 0 (la clase minoritaria) son relativamente bajas en comparación con la clase 1. Esto sugiere que el modelo podría beneficiarse de un enfoque más detallado en la detección de instancias de la clase minoritaria. No obstante, el modelo ha logrado un equilibrio sólido entre la capacidad de identificar correctamente las instancias positivas y negativas, lo que demuestra su eficacia en la tarea de clasificación en este contexto.

A fines de complementar la medición sobre la performance del modelo, en la figura 10 a continuación se grafica la curva ROC.

Figura 10. Curva ROC



Fuente: Elaboración propia en Python

Se observa que la curva ROC de la Figura 10 se aleja de la diagonal que va de (0,0) a (1,1). Esto indica que el modelo de clasificación está funcionando mejor que el azar o la suposición aleatoria. Por otra parte, cabe destacar que el modelo obtuvo un valor AUC de 0,84 lo que indica que ofrece una buena capacidad para distinguir entre las clases positivas y negativas.

Con el objetivo de identificar si el modelo está siendo justo en la manera en que realiza las predicciones se utiliza la librería *fairlearn*. El análisis de errores mediante la

librería *fairlearn* se lleva a cabo con el propósito de examinar y abordar posibles desequilibrios o sesgos presentes en los modelos de aprendizaje automático, especialmente aquellos diseñados para tareas de clasificación. Esta librería proporciona herramientas que permiten una evaluación detallada de cómo los errores de predicción se distribuyen en diferentes subgrupos de la población, especialmente en relación con atributos protegidos como género, raza u otras características demográficas sensibles.

La librería *Fairlearn* permite visualizar y cuantificar estas disparidades al proporcionar herramientas para el análisis detallado de la distribución de errores por grupos, compuesta de un panel de visualización interactivo y algoritmos de mitigación de injusticias. Se centra en mitigar los daños relacionados con la asignación y la calidad del servicio para grupos de personas basados en características como raza, género, edad o discapacidad (Bird, 2020). Esta evaluación puede ser crucial para comprender cómo el modelo está respondiendo a diferentes perfiles de individuos y para identificar áreas donde es necesario realizar ajustes para lograr un modelo más equitativo y justo. Esto contribuye a la construcción de modelos más equitativos y justos, garantizando que las decisiones basadas en estos modelos no perjudiquen a ciertos grupos de la población y promoviendo la ética en la implementación de tecnologías de inteligencia artificial.

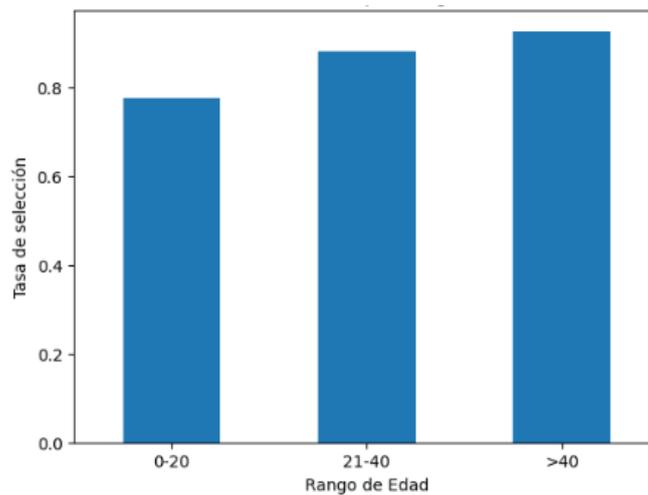
Una de las metas clave de la biblioteca *Fairlearn* es respaldar la evaluación de la equidad. La evaluación de la equidad busca responder a la pregunta: ¿Qué grupos de personas pueden verse desproporcionadamente perjudicados por un sistema de IA y de qué manera? Para ello, se evalúa cómo se desempeña el sistema para diferentes grupos poblacionales mediante el cálculo de una métrica de rendimiento en diferentes conjuntos de datos. Esto se llama evaluación desagregada (Weerts, Dudík, Edgar, Jalali, Lutz, Madaio, 2023).

Se importan entonces funciones específicas de la librería "*fairlearn.metrics*" que son esenciales para medir y entender la tasa de selección por grupo. Luego, se definen los datos que se van a analizar. Las etiquetas reales de un conjunto de prueba se almacenan en la variable *y_true*, las predicciones que hizo el modelo para ese mismo conjunto de prueba se almacenan en *y_pred* y las características sensibles, se almacenan en *sensitive_features*, que son los clústeres.

El objetivo es evaluar cuántas veces el modelo selecciona una categoría en comparación con la cantidad total de instancias en esa categoría específica. Esto se

llama *seleccion_rate*. De este modo, utilizando las funciones importadas, se calcula esta tasa de selección por grupo. Básicamente, se evalúa si el modelo tiende a seleccionar más o menos ciertas categorías dependiendo de la característica sensible. Esto es fundamental para detectar posibles sesgos o inequidades en el modelo buscando crear modelo de aprendizaje automático responsables y equitativo, permitiendo visualizar y evaluar cómo las predicciones del modelo se distribuyen en diferentes grupos. En la figura 11 a continuación se grafica la métrica *selección_rate* para 3 grupos con distintos rangos de edad de la base de clientes.

Figura 11: *Seleccion_rate* por rango de edad



Fuente: Elaboración propia en Python

Si bien se observa cierta disparidad entre el *seleccion_rate* de los distintos grupos conformados, no necesariamente indica una falta de equidad del modelo. La misma no se limita a la igualdad absoluta de las tasas de selección, sino que implica la ausencia de sesgos injustos. Los 3 grupos reflejan altos índices de selección.

Gracias al modelo configurado se logran conformar 3 nuevos clústeres según los niveles de digitalización de los clientes y se predice una tasa de recuperado estimada para cada segmento las cuales se pueden apreciar en la Tabla 3 a continuación.

Tabla 3. Tasas de recupero por clúster

Clúster Digital	Tasa de recupero estimada
Alto	96,98%
Medio	90,67%
Bajo	89,40%

Fuente: Elaboración propia en Excel

A modo de conclusión, el modelo logra pronosticar altas tasas de recuperación para todos los grupos, pero manteniendo la lógica esperada que a mayor digitalización mejor comportamiento en términos de recupero. Esto sugiere que el modelo está bastante seguro en su capacidad para predecir que los clientes en los tres grupos tienen una alta probabilidad de ser recuperados.

3.3 La predicción de comportamiento y sus beneficios en el recupero de créditos

La predicción del comportamiento en el recupero de créditos de clientes morosos representa una herramienta fundamental para las instituciones financieras y bancarias en la gestión de deudas pendientes. En esta área, científicos y expertos intentan proponer nuevos modelos de negocios financieros considerando métodos de datos masivos, en particular, métodos para el control de riesgos, análisis de mercados financieros, creación de nuevos índices de sentimiento financiero a partir de redes sociales y establecimiento de herramientas basadas en la información de diferentes formas creativas. Las técnicas de datos masivos ayudan a medir el riesgo bancario (Hasan, Popp y Oláh, 2020).

Como se desarrolló en la sección anterior, gracias al modelo configurado se logran conformar 3 nuevos clústeres según los niveles de digitalización de los clientes y se predice una tasa de recupero estimada para cada segmento. A través de esta capacidad de anticipar el comportamiento de los clientes con retrasos en sus pagos, se despliegan una serie de beneficios que contribuyen significativamente a la eficiencia y la efectividad de los procesos de recuperación.

Uno de los beneficios más destacados es la mejora radical en la estrategia general de recuperación: al prever cómo responderán los clientes morosos ante diferentes enfoques de comunicación y solución, las instituciones pueden adaptar sus acciones de manera más precisa y personalizada. Esto no solo agiliza el proceso de recuperación, sino que

también aumenta las posibilidades de éxito al enfocar los recursos en casos más propicios para el pago.

Además, esta capacidad de predicción permite una mejor asignación de recursos. Al concentrarse en los deudores con mayores probabilidades de cumplir con sus pagos, se optimiza la utilización de personal, tiempo y presupuesto, evitando destinar recursos a situaciones con probabilidades bajas de éxito. Esto conlleva a una administración más eficiente de los recursos disponibles, reduciendo costos y maximizando resultados. Al identificar tempranamente aquellos casos en los que la recuperación es poco probable, las instituciones bancarias pueden tomar decisiones informadas sobre cómo proceder. Esto evita invertir recursos valiosos en esfuerzos infructuosos y, en cambio, permite concentrarse en alternativas más viables para la recuperación.

Asimismo, es relevante también mencionar que la predicción del comportamiento contribuye a una mejora en la experiencia del cliente dado que, al diseñar estrategias de comunicación y negociación más adecuadas, se fomenta una relación transparente y menos conflictiva con los clientes en mora. Esto puede reducir la tensión entre ambas partes, mejorando la percepción del cliente hacia la institución y potencialmente evitando que la relación se deteriore aún más. En términos de mitigación de riesgos, la capacidad de anticipar incumplimientos futuros permite a las instituciones financieras adoptar medidas preventivas en la gestión del riesgo crediticio y a tomar decisiones informadas sobre la extensión de crédito en el futuro. La planificación financiera también se beneficia, ya que la predicción del comportamiento proporciona información valiosa para ajustar las proyecciones de ingresos y pérdidas.

Conclusión

A lo largo del presente trabajo, se abordó el desafío de optimizar la gestión de morosidad en una entidad bancaria local al incorporar el nivel de digitalización de los clientes. Se logró identificar patrones en el nivel de digitalización de los clientes y su relación con el comportamiento de morosidad a partir de las fuentes de datos utilizadas. Se obtuvieron nuevas segmentaciones de los clientes morosos considerando tanto las variables tradicionales como el nivel de digitalización, lo que permitirá una gestión más personalizada y eficaz. Finalmente fue posible aplicar un modelo de aprendizaje

automático para predecir el comportamiento de los nuevos segmentos de clientes y anticipar sus tasas de recupero.

En el primer capítulo se abordó literatura académica relacionada en los temas ligados al desarrollo del análisis. Este capítulo se centró en la importancia de la digitalización en la industria bancaria y su influencia en la gestión de datos, particularmente en la gestión de recuperación de morosidad. Se destacó que los bancos tienen acceso a una gran cantidad de datos sobre los clientes, lo que les permite comprender sus comportamientos financieros y diseñar estrategias más efectivas. Se exploró cómo el avance hacia el *Big Data* ha permitido a los bancos procesar y analizar estos datos para identificar patrones y tendencias en el comportamiento de los clientes. Se resaltó la necesidad de segmentar a los clientes morosos de manera más precisa, considerando factores como el nivel de deuda, el score y, en este proyecto en particular, el nivel de digitalización de los clientes.

En el segundo capítulo se llevó a cabo una optimización en la clusterización de clientes con el propósito de desarrollar estrategias de recupero más eficientes en la entidad bancario incorporando la variable de digitalización a la segmentación habitual basada en deuda y score. La extracción y preprocesamiento de datos involucraron la selección y transformación de variables relevantes para el análisis. Se realizó un análisis exploratorio y descriptivo del conjunto de datos para comprender sus características y se abordaron valores faltantes. Luego, se implementó el algoritmo *K-Means* para generar segmentos de clientes basados en similitudes, tomando en cuenta variables como actividad en canales digitales, ingreso inferido y edad. Se obtuvieron tres clústeres que reflejaron diferentes niveles de digitalización: alto, medio y bajo.

El análisis de la nueva clusterización reveló que los clientes con mayor nivel de digitalización presentan tasas de recupero más altas. De este modo, los resultados subrayan la importancia de considerar la digitalización en las estrategias de recupero ya que posibilitan una visión más precisa del comportamiento del cliente y permiten una mayor personalización de las estrategias para mejorar la eficacia de la gestión de recupero.

En el tercer y último capítulo se abordaron métodos predictivos para la previsibilidad del comportamiento de clientes morosos en la organización financiera. Se

utilizó un modelo de aprendizaje automático basado en el algoritmo *Adaboost* para predecir el comportamiento de los nuevos clústeres digitales de clientes. Se detalló el proceso de entrenamiento, optimización de hiperparámetros y evaluación de resultados. También se abordó la equidad del modelo a través de la librería "fairlearn" para identificar posibles sesgos en la predicción según diferentes subgrupos. Se concluye con que la predicción del comportamiento de clientes morosos ofrece beneficios como estrategias de recuperación personalizadas, asignación eficiente de recursos, mejora en la experiencia del cliente y mitigación de riesgos crediticios.

El trabajo realizado se considera proporciona contribuciones en términos académicos y prácticos. En primer lugar, la implementación del modelo de aprendizaje automático *Adaboost* permitió la construcción de un sistema de predicción efectivo para identificar patrones y predecir el comportamiento de clientes morosos teniendo en cuenta su nivel de digitalización. La métrica *Precision* obtenida resalta la utilidad del modelo en situaciones en las que la toma de decisiones es crucial. Las conclusiones y los enfoques aplicados en este trabajo tienen valor directo en la transferencia a contextos bancarios reales. La metodología desarrollada y las técnicas implementadas pueden ser adoptadas por instituciones financieras para optimizar la gestión de clientes morosos y mejorar la eficiencia en la recuperación de créditos

A su vez, el trabajo llevado a cabo sienta las bases para futuras investigaciones en varios aspectos. En primer lugar, la expansión del análisis a conjuntos de datos más amplios y variados puede permitir evaluar la generalización y robustez del modelo en diferentes contextos. Además, la exploración de otros algoritmos de aprendizaje automático y la comparación de sus resultados podría brindar perspectivas adicionales sobre la mejor elección de modelos para este tipo de problemas.

Este trabajo representó un esfuerzo integral y estructurado para aplicar métodos predictivos en la gestión de clientes morosos. La implementación exitosa del modelo *Adaboost*, la atención a la equidad en las predicciones y la posible transferibilidad de los resultados a contextos prácticos resalta la relevancia y la contribución a la toma de decisiones informadas y justas en el ámbito bancario. Con este enfoque, el estudio pone en relieve la importancia y el potencial de la predicción de comportamientos en el recupero de créditos, aportando un enfoque innovador para abordar este desafío.

Referencias bibliográficas

Baesens, B. (2014). "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications." Wiley.

Banco Central de la República Argentina. (2020). Comunicación "A" 7443. Ciudad Autónoma de Buenos Aires.

Banco Central de la República Argentina. (2019). Comunicación "A" 6639. Ciudad Autónoma de Buenos Aires.

Banco Central de la República Argentina. (2020). Comunicación "A" 7024. Ciudad Autónoma de Buenos Aires.

Banco Central de la República Argentina. (2016). Comunicación "A" 5598. Ciudad Autónoma de Buenos Aires

Bravo, C., Thomas, L. C., & Weber, R. (2015). Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society*, pp. 1-11. <https://doi.org/10.1057/jors.2014.50>

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., & Walker, K. (2020). Fairlearn: A toolkit for assessing and improving fairness in AI (Versión: 22 de septiembre de 2020). Microsoft, Allovus Design.

Chaurra-Gutierrez, F. A., Feregrino-Urbe, C., Perez-Sansalvador, J. C., Rodriguez-Gomez, G. (2023). QIST: One-dimensional quantum integer wavelet S-transform. *Information Sciences*, Volume 622, pp. 999-1013. <https://doi.org/10.1016/j.ins.2022.12.002>

Chen, M.-S., Han, J., & Yu, P. S. (1996). Data Mining: An Overview from a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883.

E. J. Hernández-Leal, N. D. Duque-Méndez, y J. Moreno-Cadavid (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoL.*, vol. 20, n.º 39, pp. 15–38. <http://dx.doi.org/10.22430/22565337.685>

Espino Timón, C. (2017). Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source que permiten su uso (Trabajo de

Fin de Grado). Grado en Ingeniería Informática, Business Intelligence, Universidad Oberta de Catalunya.

Espinosa, R., Jiménez, F., Palma, J. (2023). Multi-surrogate assisted multi-objective evolutionary algorithms for feature selection in regression and classification problems with time series data. *Information Sciences*, Volume 622, pp. 1064-1091. <https://doi.org/10.1016/j.ins.2022.12.004>

Fawcett, T. (2005). *An introduction to ROC analysis*. Institute for the Study of Learning and Expertise, 2164 Staunton Court, Palo Alto, CA 94306, USA.

Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. En *Machine Learning: Proceedings of the Thirteenth International Conference*. AT&T Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974-0636.

Ghosh, S., Ansari, J. (2019). Loan delinquency in banking systems: How effective are credit reporting systems? pp. 1-5. <https://doi.org/10.1016/j.ribaf.2018.07.011>

Gonzalez, N., Hoitsma, F., Knobens, A., Jastrzebska, A., Espinosa, M. L. (2023). Prolog-based agnostic explanation module for structured pattern classification. *Information Sciences*, Volume 622, pp. 1196-1227. <https://doi.org/10.1016/j.ins.2022.12.012>

Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and Big Data Mining in Banking. *Big Data and Cognitive Computing Review*, 1.

Jun Cai, Chenfu Yi. (2023). An adaptive gradient-descent-based neural networks for the on-line solution of linear time variant equations and its applications. *Information Sciences*, Volume 622, pp. 34-45. <https://doi.org/10.1016/j.ins.2022.11.157>

Mousaeirad, S. (2020). Intelligent Vector-based Customer Segmentation in the Banking Industry. pp. 5-13. <https://doi.org/10.48550/arXiv.2012.11876>

Ostapchenya, D., (2021, 11 de Junio). The Role of Big Data in Banking: How do Modern Banks Use Big Data?. Finextra. <https://www.finextra.com/blogposting/20446/the-role-of-big-data-in-banking--how-do-modern-banks-use-big-data>

Perez, R. (2021). Modelo de Scoring para la segmentación de clientes morosos usando minería de datos en una empresa de cobranzas del Perú. [Tesis de Ingeniero de Sistemas, Universidad Nacional Mayor de San Marcos]. Repositorio Institucional –

Pontificia Universidad Javeriana.

<https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16845>

Schmarzo, B. (2013). Big Data: Understanding how data powers big business. pp. 12-13. Wiley.

Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.

Taiwo, J. N., Ucheaga, E. G., Achugamonu, B. U., Adetiloye, K., Okoye, O., & Agwu, P. E. (2017). Credit Risk Management: Implications on Bank Performance and Lending Growth. Saudi Journal of Business and Management Studies, pp. 1-7. <https://doi.org/10.2139/ssrn.3122501>

Uriz, M., Paternain, D., Bustince, H., Galar, M. (2023). A supervised fuzzy measure learning algorithm for combining classifiers. Information Sciences, Volume 622, pp. 490-511. <https://doi.org/10.1016/j.ins.2022.11.161>

Verdenhofs, A., & Tambovceva, T. (2019). Evolution of customer segmentation in the era of Big Data. pp. 1-5. <https://doi.org/10.21272/mmi.2019.1-20>

Yap, B. W., Ong, S. H., & Husain, N. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Systems with applications journal, pp. 1-13. <https://doi.org/10.1016/j.eswa.2011.04.147>

Yuping, Z., Jílková, P., Guanyu, C., & Weisl, D. (2020). New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning. Advances in Economics, Business and Management Research, Volume 131, pp. 1-7. <https://doi.org/10.2991/aebmr.k.200324.170>

Zhenlong Man, Jinqing Li, Xiaoqiang Di, Ripei Zhang, Xusheng Li, Xiaohan Sun. (2023). Research on cloud data encryption algorithm based on bidirectional activation neural network. Information Sciences, Volume 622, pp. 629-651. <https://doi.org/10.1016/j.ins.2022.11.089>

Trabajo Final de Especialización de Mariano Esmail:

“La digitalización bancaria como variable relevante para la gestión de deuda y segmentación de clientes morosos.

Implementación de modelos de Machine Learning”

Mentora: Mg. Natalia Salaberry

Fecha: septiembre 2023

El alumno Mariano Esmail en su trabajo final de especialización logra desarrollar de manera concisa y acabada su proyecto inicial de trabajo. Plantea de manera clara la existencia de una oportunidad de mejora en una organización que luego resuelve con las herramientas aprendidas en el campo disciplinar de la Especialización en Métodos Cuantitativos para la Gestión y Análisis de Datos en Organizaciones. En este sentido, el problema propuesto a resolver se vincula a la determinación de una nueva segmentación de cartera de clientes morosos de una organización bancaria a partir del nivel de digitalización que poseen como variable innovadora y relevante, contribuyendo a que la organización pueda rediseñar la estrategia de recupero de morosos de manera más asertiva.

Para cumplimentar con el objetivo planteado, en un primer capítulo desarrolla una clara contextualización de las implicancias de la gestión de datos en organizaciones bancarias. A partir de este punto, describe la situacionalidad en base a las oportunidades y desafíos que presenta la disponibilidad de grandes volúmenes de datos. Pero también advierte sobre la necesidad e importancia de una gestión adecuada de estos para lograr la construcción de valor agregado para la toma de decisiones.

En un segundo capítulo, se centra en exponer de manera concisa los datos con los cuales trabajará, así como las herramientas metodológicas a utilizar y que fueron aprendidas en diversas materias de la especialización. Desde el aspecto técnico de implementación de modelos seleccionados, realiza un trabajo perfecto, respetando todas las etapas de aplicación, análisis y evaluación de resultados. Supo seleccionar adecuadamente los modelos acordes que le permitieron cumplimentar con el objetivo buscado.

Finalmente, en un tercer capítulo expone como los resultados obtenidos permiten generar valor agregado para la toma de decisiones a partir de su análisis completo. La evaluación y análisis de resultados obtenidos es lo que le permite identificar la relevancia del nivel de digitalización de los clientes morosos cumplimentando con su objetivo principal. Esto, es realizado en completo detalle.

De esta manera, el alumno logra cumplimentar con el objetivo principal y los específicos de forma ordenada y clara, realizando un trabajo completamente estructurado en el marco de las pautas establecidas. Existe una coherencia entre la problemática planteada, el título y las palabras claves lo que le permitió poder elaborar un planteo y desarrollo adecuado.